Title: PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair**

Authors: Takashi Ochi[1,6], Andrew N. Blackford[2,6], Julia Coates[2], Satpal Jhujh[2], Shahid Mehmood[3], Naoka Tamura[4], Jon Travers[2], Qian Wu[1], Viji M. Draviam[4], Carol V. Robinson[3], Tom L. Blundell[1,*], Stephen P. Jackson[1,2,5,*]

Affiliations:

[1]Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK.

[2]Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK.

[3]Physical and Theoretical Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QZ, UK.

[4]Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK.

[5]Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK.

[6]These authors contributed equally.

*Correspondence to: s.jackson@gurdon.cam.ac.uk or tlb20@cam.ac.uk

**This manuscript has been accepted for publication in Science. This version has not undergone final editing. Please refer to the complete version of record at http://www.sciencemag.org/.

Abstract:

XRCC4 and XLF are two structurally-related proteins that function in DNA double-strand break (DSB) repair. Here, we identify human PAXX (PAralog of XRCC4 and XLF; also called C9orf142) as a new XRCC4-superfamily member, and show that its crystal structure resembles that of XRCC4. PAXX interacts directly with the DSB-repair protein Ku and is recruited to DNA-damage sites in cells. Using RNA interference and CRISPR-Cas9 to generate $PAXX^{-/-}$ cells, we demonstrate that PAXX functions with XRCC4 and XLF to mediate DSB repair and cell survival in response to DSB-inducing agents. Finally, we reveal that PAXX promotes Ku-dependent DNA ligation *in vitro*, and assembly of core non-homologous end-joining (NHEJ) factors on damaged chromatin in cells. These findings identify PAXX as a new component of the NHEJ machinery.

One Sentence Summary: PAXX, a paralog of XRCC4, promotes DNA double-strand break repair.

Main Text:

DNA double-strand breaks (DSBs) are toxic lesions that arise in cells exposed to agents such as ionizing radiation (IR), and are also generated as intermediates during V(D)J (variable, diversity, joining) and class-switch recombination at immune-receptor gene loci (*1*). If unrepaired or repaired incorrectly, DSBs cause cell death or genome instability, and defects in DSB repair components cause hereditary disorders with symptoms including immunodeficiency, neurodegeneration, infertility and/or increased cancer predisposition (*2*). A key DNA repair pathway is non-homologous end-joining (NHEJ), which involves initial recognition of a DSB by the Ku70-Ku80 heterodimer followed by the assembly of additional factors including the DNA-dependent protein kinase catalytic subunit (DNA-PKcs), X-ray cross-complementing protein 4 (XRCC4) and XRCC4-like factor (XLF; also called Cernunnos), with XRCC4 playing a prime role in recruiting DNA ligase IV (LIG4) to carry out the DSB-joining reaction (*3*).

XRCC4 and XLF, together with spindle-assembly abnormal protein 6 (SAS6), comprise a homologous superfamily of structurally-related proteins (*4*). To identify other members of this protein superfamily, we used a bioinformatics approach, which suggested that the 22-kDa human protein, C9orf142, could be an XRCC4 paralog (Table S1 and S2). While displaying little overall primary sequence similarity, sequence alignments indicated that the N-terminal portion of C9orf142 contains the present-in-SAS6 (PISA) motif that is conserved throughout the XRCC4 superfamily (*5*), and shares a conserved tryptophan in this motif with XRCC4 and XLF (Fig. 1A and Fig. S1A). We have therefore named this previously uncharacterized protein PAXX (PAralog of XRCC4 and XLF). While having a wide evolutionary distribution (Fig. S1B), we could not identify PAXX orthologs in insects or fungi.

We solved crystal structures of PAXX residues 1-145, 1-166 and the full-length protein ($PAXX^{1-145}$, $PAXX^{1-166}$ and $PAXX^{1-204}$) at resolutions of 2.46, 2.35 and 3.45 Å respectively (Table S3). PAXX residues 1-113 form a head domain that is structurally closely related to those of XRCC4, XLF and SAS6 (*6–11*) (Fig. 1, A and B). This is followed by a 31-amino-acid α-helix, forming a coiled-coil with the other protomer to make a PAXX homodimer (in the asymmetric unit) in a similar manner to the dimeric interfaces formed by XRCC4, XLF and SAS6. Contacts within the PAXX crystal indicate that the protein has potential to form higher-order protein filaments with similarities to those of its paralogs (Fig. S2A), in which two β-sheets, each comprising strands β5-β7 from different PAXX dimers, form a β-sandwich around a dyad axis running between the sheets and orthogonally to the strands (Fig. 1C). However, small angle X-ray scattering (SAXS) data of $PAXX^{1-145}$ in solution are well explained by the scattering curve calculated from the structure of the dimer of the construct (Fig. S2B), and electrospray mass spectroscopy confirmed that PAXX predominantly forms dimers in solution under the conditions used (Fig. 1D and Table S4). These data thus suggest that the preferred native state for PAXX is a dimer. While our analyses indicate that the conformation of PAXX is distinct from those of other XRCC4-superfamily members, its overall structural properties most resemble those of XRCC4 (Fig. 1E and S2C).

To gain insights into PAXX function, we expressed green-fluorescent protein (GFP), GFP-tagged wild-type (WT) PAXX and GFP-tagged C-terminally truncated $PAXX^{1-145}$ in human 293FT cells, purified these and identified potential binding partners by mass spectrometry. This revealed that the only proteins that bound GFP-$PAXX^{WT}$ but not GFP-$PAXX^{1-145}$ or GFP corresponded to Ku70 and Ku80. Immunoprecipitation and western blot analyses confirmed this interaction and established that endogenous PAXX and Ku interact (Fig. 2, A and B). Also, in

reciprocal experiments Ku70 (GFP-tagged at its endogenous gene locus) interacted with PAXX (Fig. S3A). Given that GFP-PAXX$^{1-145}$ did not interact with Ku and because the extreme C-terminus of PAXX has been highly conserved through evolution (Fig. 2C), we speculated that PAXX C-terminal residues might mediate Ku binding. To test this, we synthesized a biotinylated peptide encompassing PAXX residues 177-204 and found that it specifically retrieved two major proteins that were identified by mass spectrometry to be Ku70 and Ku80 (Fig. 2D); this was confirmed by western blotting (Fig. S3B). Furthermore, mutating two of the most highly conserved residues in the PAXX C-terminus (V199 and F201) to alanine in GFP-PAXX^{WT} or biotinylated PAXX$^{177-204}$ abolished interaction with Ku (Fig. 2E and S3B). In line with these findings, addition of PAXX$^{177-204}$ peptide to cell lysates inhibited the interaction of PAXX with Ku (Fig. S3C), and surface-plasmon-resonance (SPR) studies with purified proteins established that DNA-bound Ku and PAXX^{WT} interacted directly, while PAXX$^{V199A/F201A}$ did not bind to Ku-DNA detectably (Fig. S3, D and E). Thus, PAXX binding to Ku-DNA is direct and is mediated by the PAXX C-terminus.

In vivo, PAXX was predominantly nuclear (Fig. S4, A and B), and GFP-PAXX localized to DNA damage generated by laser micro-irradiation of live cells (Fig. 3A). To test if PAXX might be involved in DSB repair by NHEJ, we depleted human U2OS cells of PAXX using multiple small-interfering RNAs (siRNAs; Fig. S4C), then performed clonogenic survival assays after exposing the cells to IR. Cells in which PAXX was depleted were significantly more radiosensitive than control cells, displaying similar sensitivities to cells depleted of XRCC4 (Fig. 3B and S4D). Furthermore, expression of PAXX^{WT} but not PAXX$^{V199A/F201A}$ restored IR resistance in PAXX-depleted cells (Fig. 3B). As PAXX$^{V199A/F201A}$ is impaired for Ku binding,

these data support a model in which the PAXX-Ku interaction is crucial for PAXX function in DNA repair.

To verify and extend the above conclusions, we used CRISPR-Cas9 gene-editing (*12*) in non-transformed human RPE-1 cells to generate $PAXX^{-/-}$ clones (Fig. S5). Like siRNA-treated U2OS cells, $PAXX^{-/-}$ cells were hypersensitive to IR and the radiomimetic drug phleomycin (Fig. 3C and Fig. S6B). Furthermore, by depleting XRCC4 or XLF in $PAXX^{+/+}$ or $PAXX^{-/-}$ cells (Fig. S6C), we established that combined loss of PAXX and XRCC4, or PAXX and XLF, did not cause IR sensitivity greater than that of $PAXX^{-/-}$ cells or cells depleted of XRCC4 or XLF alone (Fig. 3C and Fig. S6D), implying that PAXX functions epistatically with XRCC4 and XLF to promote IR resistance *via* classical NHEJ. Given that NHEJ-deficient cells display defective resolution of IR-induced γH2AX foci, we compared the appearance and disappearance of these foci by immunofluorescence microscopy in $PAXX^{+/+}$ and $PAXX^{-/-}$ cells. While PAXX loss did not impair γH2AX focus formation, we observed a significant defect in γH2AX focus resolution in multiple $PAXX^{-/-}$ clones (Fig. 3D). Furthermore, both $PAXX^{-/-}$ RPE-1 and PAXX-depleted U2OS cells were impaired in repairing DSBs as measured by neutral comet assays (Fig. 3E and Fig. S6E), and PAXX-depleted U2OS cells were also defective in random-plasmid integration, which occurs through NHEJ events (Fig. S6F) (*13*). As with other NHEJ factors such as XRCC4, PAXX loss did not impair checkpoint signaling (Fig. S7, A and B).

Subsequent biochemical investigations established that, while not binding DNA detectably on its own, PAXX retarded the electrophoretic mobilities of DNA complexes containing two Ku molecules (Fig. 4A and Fig. S8A-C). Furthermore, such binding was abrogated in the presence of a large excess of $PAXX^{177-204}$ peptide, or when the extreme PAXX C-terminal region was

absent or contained alanine substitutions of V199 and F201 (Fig. 4A and S8D). We next tested whether PAXX affected DNA ligation by LIG4 *in vitro* in a manner dependent on its ability to bind Ku. Indeed, $PAXX^{WT}$ but not $PAXX^{V199A/F201A}$ markedly stimulated double-stranded DNA ligation in reactions containing the XRCC4/LIG4 complex, but only in the presence of Ku (Fig. 4B). We speculated that PAXX might act as a scaffold to stabilize two Ku heterodimers at DNA ends and thus promote assembly and/or stability of the NHEJ machinery at DSB sites. To test this, we treated $PAXX^{+/+}$ and $PAXX^{-/-}$ cells with phleomycin and examined the association of NHEJ proteins with chromatin by western blotting. This revealed that PAXX deficiency produced substantial defects in the ability of Ku, DNA-PKcs, XRCC4 and XLF to assemble on chromatin in response to DNA damage, without affecting the overall levels of these proteins (Fig. 4C and S8E).

In conclusion, we have identified and characterized a novel XRCC4-superfamily member, PAXX. We have shown that PAXX binds Ku and promotes DSB repair at the biochemical and cellular levels and stabilizes NHEJ-protein assembly at DSB sites (Fig. 4D), thus establishing PAXX as a hitherto uncharacterized NHEJ factor.

References:

1. S. P. Jackson, J. Bartek, The DNA-damage response in human biology and disease, *Nature* **461**, 1071–1078 (2009).

2. L. Woodbine, A. Gennery, P. Jeggo, The clinical impact of deficiency in DNA non-homologous end-joining, *DNA Repair* **16**, 84–96 (2014).

3. G. Williams *et al.*, Structural insights into NHEJ: Building up an integrated picture of the dynamic DSB repair super complex, one component and interaction at a time, *DNA Repair* **17**, 110–120 (2014).

4. T. Ochi, Q. Wu, T. Blundell, The spatial organization of non-homologous end joining: From bridging to end joining, *DNA Repair* **17**, 98–109 (2014).

5. S. Leidel, M. Delattre, L. Cerutti, K. Baumer, P. Gonczy, SAS-6 defines a protein family required for centrosome duplication in *C. elegans* and in human cells, *Nat. Cell Biol.* **7**, 115–125 (2005).

6. M. Junop *et al.*, Crystal structure of the Xrcc4 DNA repair protein and implications for end joining, *EMBO J.* **19**, 5962–5970 (2000).

7. B. Sibanda *et al.*, Crystal structure of an Xrcc4-DNA ligase IV complex, *Nat. Struct. Biol.* **8**, 1015–1019 (2001).

8. S. N. Andres, M. Modesti, C. J. Tsai, G. Chu, M. S. Junop, Crystal Structure of Human XLF: A Twist in Nonhomologous DNA End-Joining, *Mol. Cell* **28**, 1093–1101 (2007).

9. Y. Li *et al.*, Crystal structure of human XLF/Cernunnos reveals unexpected differences from XRCC4 with implications for NHEJ, *EMBO J.* **27**, 290–300 (2008).

10. M. van Breugel *et al.*, Structures of SAS-6 Suggest Its Organization in Centrioles, *Science* **331**, 1196–1199 (2011).

11. D. Kitagawa *et al.*, Structural basis of the 9-fold symmetry of centrioles, *Cell* **144**, 364–375 (2011).

12. P. D. Hsu, E. S. Lander, F. Zhang, Development and applications of CRISPR-Cas9 for genome engineering, *Cell* **157**, 1262–1278 (2014).

13. P. Ahnesorg, P. Smith, S. P. Jackson, XLF Interacts with the XRCC4-DNA Ligase IV Complex to Promote DNA Nonhomologous End-Joining, *Cell* **124**, 301–313 (2006).

Acknowledgments:

T.O. and T.L.B. are supported by the Wellcome Trust. The Jackson lab is funded by Cancer Research UK (CRUK) program grant C6/A11224, the European Research Council and the European Community Seventh Framework Programme grant agreement no. HEALTH-F2-2010-259893 (DDResponse). Core infrastructure funding to the Jackson lab is provided by CRUK (C6946/A14492) and the Wellcome Trust (WT092096). S.P.J. receives his salary from the University of Cambridge, supplemented by CRUK. V.M.D. is a CRUK Career Development

Fellow. The Draviam lab is funded by a CRUK CDA (C28598/A9787). We thank M. Hyvönen, V. Bolanos-Garcia and L. Hanakahi for reagents, K. Scott for her assistance with SPR, beamline scientists at I03 and I22 for their help at the Diamond Light Source, and J. Brown, K. Inoue, Y. Kimata, M. Lamers, B. Luisi, N. Lukashchuk, R. Nishi, C. le Sage, C. Schmidt and P. Wijnhoven for useful discussions and technical assistance. Crystallization and initial X-ray diffraction experiments were performed in the X-ray crystallographic facility at the Department of Biochemistry, University of Cambridge, with help from the Facility Manager, D. Chirgadze. PDB accession numbers of $PAXX^{1-166}$ and $PAXX^{1-204}$ are 3WTD and 3WTF respectively.

Figure legends:

Fig. 1. Crystal structure of PAXX. (**A**) Domain architecture of human PAXX and other XRCC4-superfamily members. Sequence identities between human PAXX and XRCC4, XLF and SAS6 are 10.9, 11.2 and 10.1% respectively. (**B**) Structure of a PAXX dimer with the two polypeptide chains shown in cyan and pink. The N- and C-termini of the structure are indicated as N-ter and C-ter respectively. (**C**) Residues mediating the PAXX dimerization interface, with β-sheet sandwich-like packing between protomers in adjacent asymmetric units. The packing of each strand from different protomers is shown in surface and stick representations. Black dotted lines are hydrogen bonds between side-chain pairs (S95 and S99, T97 and T97, S99 and S95) and between the side-chain of D108 of each protomer with the main chain of A104 of the adjacent protomer in the crystals. (**D**) ES-MS profile showing that $PAXX^{1-204}$ is a dimer. Three charge states are observed for the dimer. The main charge state 13+ is labeled in the mass spectrum. **(E)** Comparison of XRCC4-superfamily members. The head domains of PAXX (cyan), XRCC4 (silver), XLF (magenta) and SAS6 (lime) are superimposed.

Fig. 2. The PAXX C-terminus interacts with Ku. (**A**) GFP-pulldowns showing that GFP-$PAXX^{WT}$ but not GFP-$PAXX^{1-145}$ transiently overexpressed in 293FT cells interacts with Ku. (**B**) Co-IP from HeLa nuclear extracts showing that endogenous PAXX and Ku interact. (**C**) Sequence alignment of the C-termini of PAXX orthologs. Conserved residues are indicated with reverse shading and similar residues are highlighted in grey. (**D**) Peptide pulldowns from HeLa nuclear extracts using control (H2AX) or $PAXX^{177-204}$ peptides analysed by silver staining. "M" represents protein markers, and numbers represent molecular weights in kDa. (**E**) GFP-pulldown showing that PAXX residues V199 and F201 are required for Ku binding in the context of full-length PAXX.

Fig. 3. PAXX is required for DSB repair in human cells. (**A**) GFP-tagged PAXX accumulates at sites of laser micro-irradiation in U2OS cells. White arrowheads indicate the path of the laser used to induce DSBs. (**B**) Clonogenic survival assay showing that PAXX depletion in U2OS cells causes radiosensitivity and that this is rescued by exogenous expression of $PAXX^{WT}$ but not $PAXX^{V199A/F201A}$. In this experiment and those below, error bars represent the standard error of the mean (SEM) from 3 independent experiments. (**C**) Clonogenic survival assay showing that $PAXX^{-/-}$ cells are radiosensitive and that PAXX loss is epistatic with XRCC4 depletion. (**D**) $PAXX^{-/-}$ cells display persistent γ-H2AX foci after IR. Cells with >5 foci were scored as positive, and were co-stained with Cyclin A to eliminate S and G2 cells from analysis. At least 100 cells were scored per condition. "cl." indicates clone number. (**E**) PAXX is required for cellular DSB repair as measured by neutral comet assay. R/D ratios represent mean tail length of cells treated

with 40 μg/ml phleomycin for 2 hours and allowed to recover (R) for 2 hours over mean tail length of cells damaged (D) for 2 hours without recovery.

Fig. 4. PAXX promotes NHEJ *in vitro* and stabilizes NHEJ proteins on damaged chromatin. (**A**) EMSA of Ku and PAXX derivatives with 50-bp 6-FAM(6-carboxyfluorescein)-labelled DNA. 200 nM PAXX and 20 nM Ku were added where indicated. (**B**) Stimulation of DNA-end ligation by PAXX. 50 ng of pcDNA3.1(-) digested by *Eco*RV was incubated with XRCC4/LIG4 (25 nM), Ku (25 nM), PAXX (250 nM) and $PAXX^{V199A/F201A}$ (250 nM) as indicated (left). Ligation efficiency (right) was calculated as a percentage of ligated plasmid from four independent assays. (**C**) Chromatin fractionation of $PAXX^{+/+}$ and $PAXX^{-/-}$ cells treated with phleomycin as indicated. Note that DNA-damage-dependent chromatin recruitment of proteins such as RPA, that function in DNA repair pathways other than NHEJ, were unaffected by PAXX loss. (**D**) Model of PAXX in NHEJ. Two Ku-bound DNA-ends are bound by a PAXX dimer *via* its C-termini, which stabilizes the NHEJ machinery to promote DNA-end ligation.

Supplementary Materials:

Materials and Methods

Figs. S1-S9

Tables S1-S4

References (14-39)

Figure 1

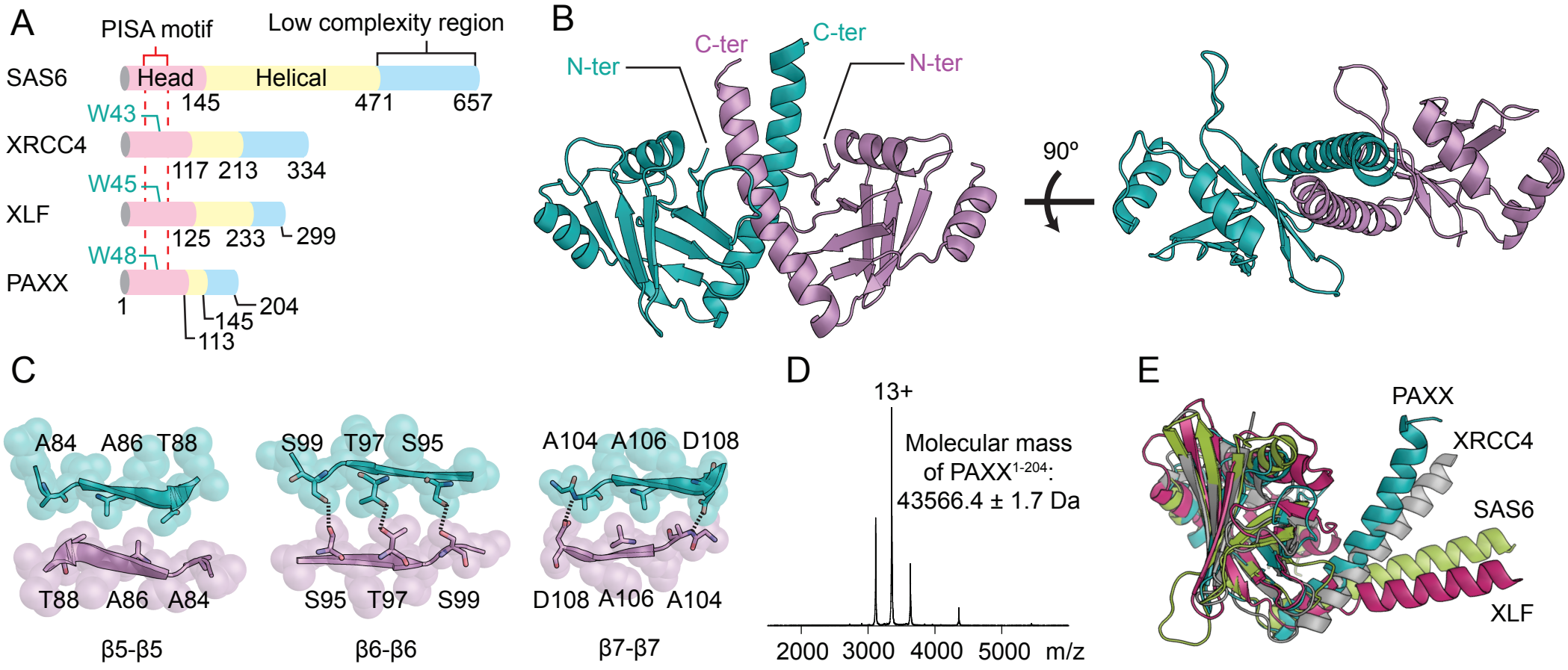

A
PISA motif
Low complexity region
SAS6
Head
Helical
145
471
657
W43
XRCC4
117
213
334
W45
XLF
125
233
299
W48
PAXX
1
145
204
113
B
N-ter
C-ter
C-ter
N-ter
90°
C
A84
A86
T88
T88
A86
A84
S99
T97
S95
S95
T97
S99
A104
A106
D108
D108
A106
A104
β5-β5
β6-β6
β7-β7
D
13+
Molecular mass
of PAXX1-204:
43566.4 ± 1.7 Da
2000
3000
4000
5000
m/z
E
PAXX
XRCC4
SAS6
XLF

Figure 2

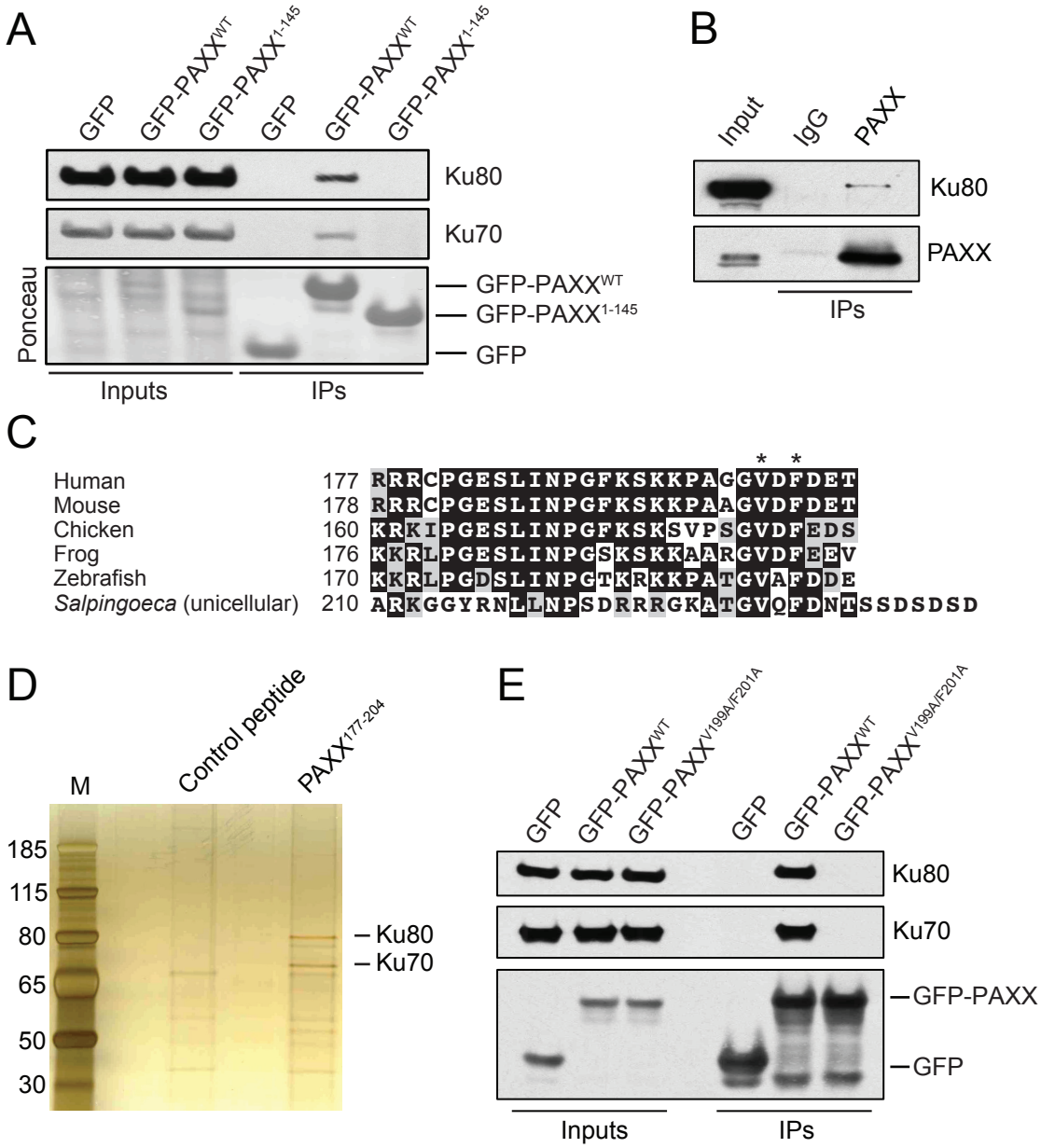

A
GFP
GFP-PAXX^WT
GFP-PAXX^1-145
GFP
GFP-PAXX^WT
GFP-PAXX^1-145
Ku80
Ku70
Ponceau
GFP-PAXX^WT
GFP-PAXX^1-145
GFP
Inputs
IPs
B
Input
IgG
PAXX
Ku80
PAXX
IPs
C
Human 177 RRRCPGESLINPGFKSKKPAGGVDFDET
Mouse 178 RRRCPGESLINPGFKSKKPAAGVDFDET
Chicken 160 KRKIPGESLINPGFKSKSVPSGVDFEDS
Frog 176 KKRLPGESLINPGSKSKKAARGVDFEEV
Zebrafish 170 KKRLPGDSLINPGTKRKKPATGVAFDDE
Salpingoeca (unicellular) 210 ARKGGYRNLLNPSDRRRGKATGVQFDNTSSDSDSD
D
M
Control peptide
PAXX^177-204
185
115
80
65
50
30
Ku80
Ku70
E
GFP
GFP-PAXX^WT
GFP-PAXX^V199A/F201A
GFP
GFP-PAXX^WT
GFP-PAXX^V199A/F201A
Ku80
Ku70
GFP-PAXX
GFP
Inputs
IPs

Figure 3

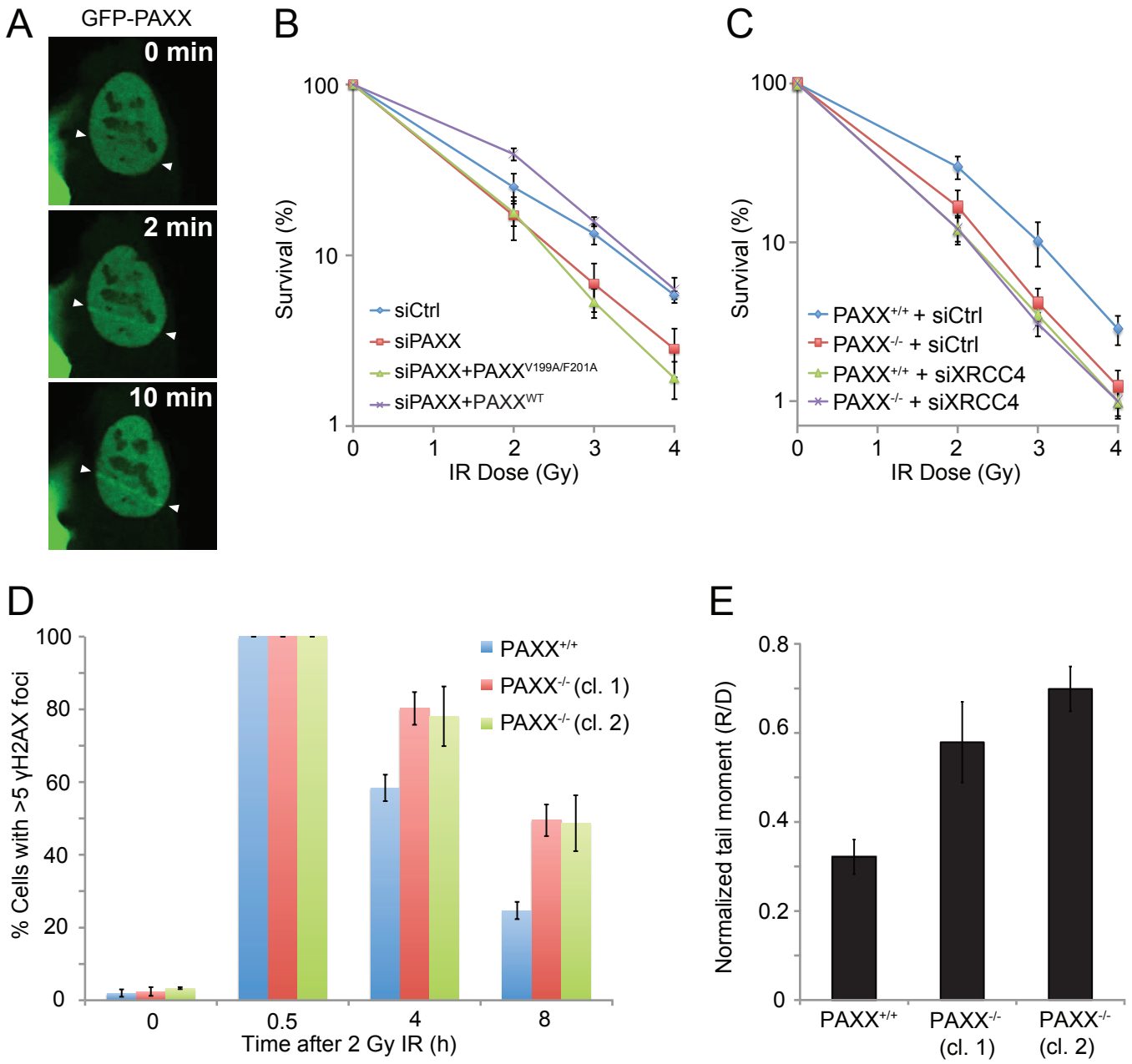
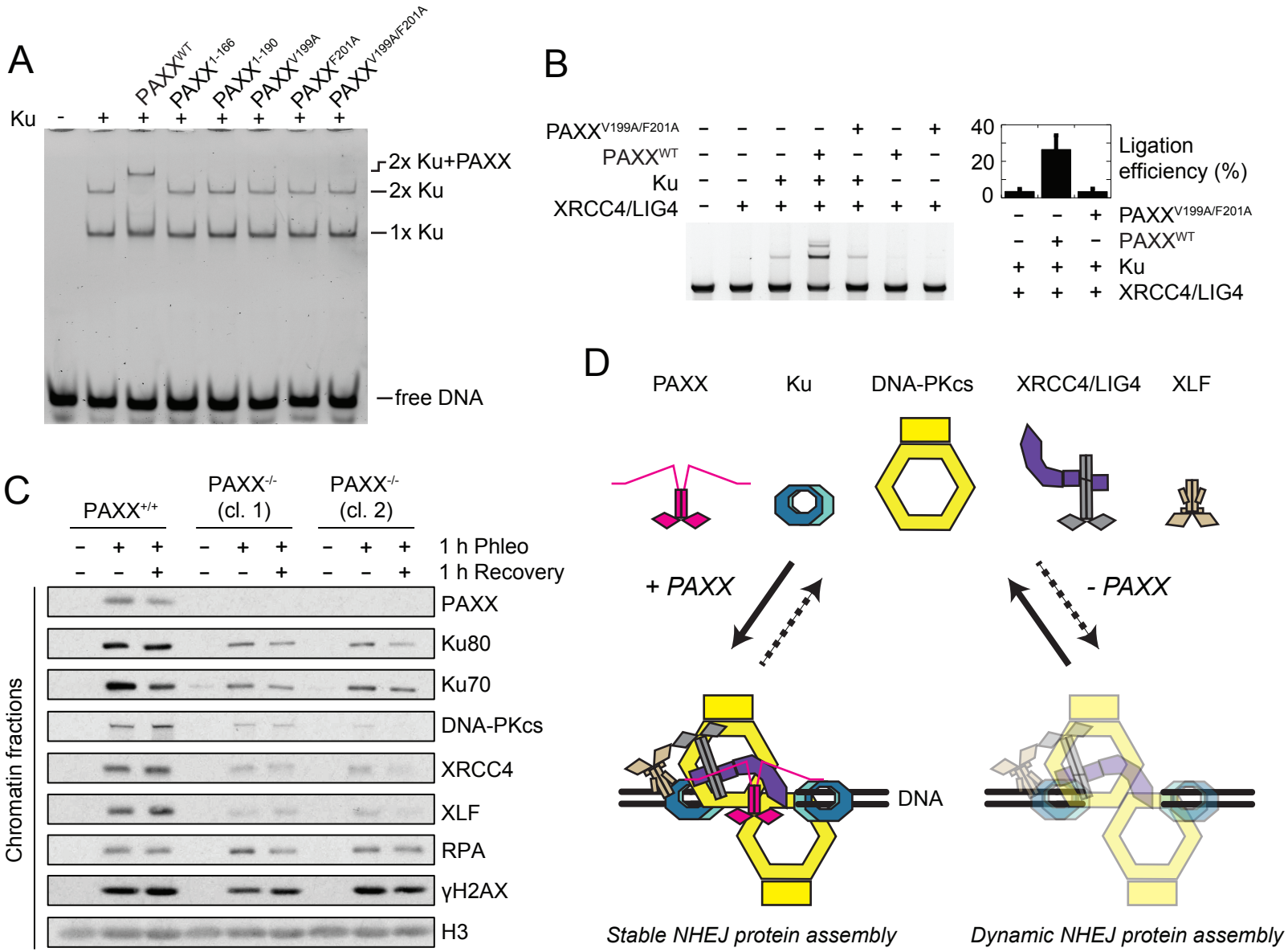

A
GFP-PAXX
0 min
2 min
10 min
B
Survival (%)
IR Dose (Gy)
siCtrl
siPAXX
siPAXX+PAXX^V199A/F201A
siPAXX+PAXX^WT
C
Survival (%)
IR Dose (Gy)
PAXX+/+ + siCtrl
PAXX-/- + siCtrl
PAXX+/+ + siXRCC4
PAXX-/- + siXRCC4
D
% Cells with >5 γH2AX foci
Time after 2 Gy IR (h)
PAXX+/+
PAXX-/- (cl. 1)
PAXX-/- (cl. 2)
E
Normalized tail moment (R/D)
PAXX+/+
PAXX-/- (cl. 1)
PAXX-/- (cl. 2)

Figure 4


A
PAXX^WT
PAXX^1-166
PAXX^1-190
PAXX^V199A
PAXX^F201A
PAXX^V199A/F201A
Ku
2x Ku+PAXX
2x Ku
1x Ku
free DNA
B
PAXX^V199A/F201A
PAXX^WT
Ku
XRCC4/LIG4
Ligation efficiency (%)
C
PAXX+/+
PAXX-/- (cl. 1)
PAXX-/- (cl. 2)
1 h Phleo
1 h Recovery
Chromatin fractions
PAXX
Ku80
Ku70
DNA-PKcs
XRCC4
XLF
RPA
γH2AX
H3
D
PAXX
Ku
DNA-PKcs
XRCC4/LIG4
XLF
+ PAXX
- PAXX
DNA
Stable NHEJ protein assembly
Dynamic NHEJ protein assembly

Supplementary Materials for

PAXX, A Paralog Of XRCC4 And XLF, Interacts With Ku To Promote DNA Double-Strand Break Repair

Takashi Ochi, Andrew N. Blackford, Julia Coates, Satpal Jhujh, Shahid Mehmood, Naoka Tamura, Jon Travers, Qian Wu, Viji M. Draviam, Carol V. Robinson, Tom L. Blundell, Stephen P. Jackson

correspondence to: s.jackson@gurdon.cam.ac.uk or tlb20@cam.ac.uk

This PDF file includes:

Materials and Methods

Construct design and purification

The synthesised cDNA of human *PAXX* (*C9orf142*) optimised for *Escherichia coli* expression (Life Technologies) was cloned into the pHAT4 vector (*14*), with a hexa-histidine tag followed by a TEV cleavage site used for $PAXX^{1-145}$, $PAXX^{1-166}$ and $PAXX^{1-204}$, or into pGAT3 vector (a gift from Dr. M. Hyvönen), with a hexa-histidine tag followed by a GST tag and a TEV cleavage site used for $PAXX^{1-190}$ and $PAXX^{1-204}$. The pGAT3 vector carrying the $PAXX^{1-204}$ open-reading frame (ORF) was used for site-directed mutagenesis to produce $PAXX^{V199A}$, $PAXX^{F201A}$ and $PAXX^{V199A/F201A}$. The coding regions of $PAXX^{1-204}$ and $PAXX^{V199A/F201A}$ were amplified from the vectors by PCR and cloned into the pcDNA3.1(-) vector (a gift from Dr. V. Bolanos-Garcia) together with GFP or FLAG tags using In-Fusion (Clontech). Vectors were sequenced by the DNA-sequencing facility, Department of Biochemistry, University of Cambridge. Bacterial expression vectors were transformed into BL21(DE3) (New England Biolabs). PAXX derivatives with 6xHis tags only were purified by Ni-affinity column and gel filtration. $PAXX^{1-166}$ and $PAXX^{1-204}$ were further purified by cation-exchange chromatography. Derivatives with GST tags were purified by GST-affinity column, Ni-affinity column and gel filtration chromatography. All tags were cleaved by TEV after the first chromatography step. Proteins were stored in final buffer 20 mM Tris-HCl pH 7.5, 100 mM NaCl, 2 mM DTT) at -80 ºC. Bands from the final SDS-PAGE gels were analyzed using mass spectroscopy by Dr. L. Packman (PNAC facility, Department of Biochemistry, University of Cambridge) or Dr. B. Kessler (Target Discovery Institute, University of Oxford). Purification of XRCC4/LIG4 was carried out as described previously (*15*). The co-expression vector of Ku was a generous gift from Dr. L. Hanakahi (Johns Hopkins University), and protein expression and purification were carried out as described (*16*). However, the lysis step was modified and an additional gel filtration step was added at the end. For lysis, wet cell pellet was suspended in 5 ml/g of cells in 50 mM sodium phosphate buffer pH 7.5, 150 mM NaCl, 5% (v/v) glycerol, 2 mM 2-mercaptoethanol, 10 μg/ml of RNase, 20 μg/ml of DNase supplied with cOmplete,

EDTA-free protease inhibitor cocktail (Roche). Lysate was incubated at 4 ºC at 30 min after the re-suspended cells were sonicated. Lysate was diluted 5-fold to make the final lysis buffer (50 mM sodium phosphate buffer pH 8.0, 1 M NaCl, 10% (v/v) glycerol, 400 mM ammonium acetate, 2 mM 2-mercaptoethanol, 10 mM imidazole, protease inhibitor cocktail). Debris was removed by centrifuging at 25,000 x g for 30 min, 4 ºC. For gel filtration, Ku eluted from 5 ml HiTrap heparin HP column (GE Healthcare) was concentrated and loaded onto Superdex200 16/60 or 10/300 (GE Healthcare), which was equilibrated with 20 mM HEPES pH 7.5, 500 mM NaCl, 5% (v/v) glycerol, 2 mM DTT, depending on sample volumes. A single peak containing Ku heterodimer was collected and diluted 5-fold with 20 mM HEPES pH 7.5, 100 mM NaCl, 5% (v/v) glycerol, 2 mM DTT. The sample was concentrated, analyzed by SDS-PAGE (Fig. S8A), and stored at -80 ºC after being snap-frozen in liquid nitrogen.

Crystallisation and structure determination

10 mg/ml of a) PAXX$^{1-145}$, b) PAXX$^{1-166}$ and c) PAXX$^{1-204}$ were crystallized at 18 ºC in the following conditions respectively: a) 1.4-1.6 M ammonium sulphate, 100 mM Tris-HCl pH 7.6-8.4; b) 1.5 M ammonium sulphate, 12-15% (v/v) glycerol, 100 mM Tris-HCl pH 8.2-8.8; and c) 1 M sodium malonate pH 6.0, 1% (v/v) 1,4-dioxane. The derivative crystals of PAXX$^{1-145}$ were prepared by soaking crystals in the crystallization solution with 1 mM $KAu(CN)_2$ for 120 min and by back-soaking in the crystallization solution for 120 min. Native and derivative crystals were then cryo-protected in crystallization solutions containing 30% (v/v) glycerol before flash freezing in liquid nitrogen. X-ray diffraction data were collected at I03 in the Diamond Light Source (Oxford, UK). Collected data were processed by XDS (*17*) and scaled using Aimless (*18*) (Table S3). The phenix.autosol module of PHENIX suite (*19*) was used to calculate experimental phases for structure factors by SAD method (FOM: 0.311 and overall score: 49.85). The initial model was built using the phenix.autobuilt module and used to solve the structure of PAXX$^{1-166}$ by molecular replacement using the phenix.phaser module because the diffraction data quality of PAXX$^{1-166}$ was better than that of PAXX$^{1-145}$ (Table S3). The placed model was manually and computationally refined using Coot (*20*) and the phenix.refine module against the native data set of PAXX$^{1-166}$ until the improvement of

the model converged (Table S3). The structural model was used to calculate phases of structure factors of $PAXX^{1-204}$ datasets by molecular replacement using the phenix.phaser module. Model building and refinement were performed in a similar manner to those of $PAXX^{1-166}$ (Table S3). The constructs crystallized in the same space group with similar cell parameters (Table S3) and with very similar crystal packing. Due to their being disordered, PAXX residues after 145 were not observed in the crystal structures.

Electrospray mass spectroscopy

Prior to MS analysis, samples were buffer exchanged into 200 mM ammonium acetate (pH 7.8) using micro biospin 6 columns (BioRad). Mass spectra were obtained on a Q-ToF mass spectrometer previously modified for high-mass transmission (*21*). The samples were introduced using a method to preserve noncovalent interactions (*22*). The mass spectra were obtained under following instrumental conditions: sample cone 100 V, capillary voltage 1500 V, extractor cone 10 V, collision cell energy 10 V, source backing pressure 4.5 x 10^{-3} mbar. Measured molecular weights of $PAXX^{1-145}$ and $PAXX^{1-204}$ are shown in Table S4.

Small angle X-ray scattering (SAXS)

SAXS data of 2-10 mg/ml of $PAXX^{1-145}$ were collected at I22 of Diamond Light Source. Data collection was carried out at 5 ºC using a BIOSAXS robot. Buffer (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 2 mM DTT) and each sample was exposed to the beam alternately in 20 frames of 1 s. Scattering intensities were measured using the Pilatus 2M detector at 3.2 m distance. Data reduction was carried out using DAWN. ScÅtter (http://www.bioisis.net/) was used to merge 2 and 10 mg/ml scattering data of the construct. The samples were shown to be monodisperse by checking Guinier plots in the Guinier region generated using ScÅtter (Figure S9A). The distance distribution function (Figure S9B) and *Rg* (Table S4) of the construct was calculated from the merged data using GNOM (*23*). The complete model of $PAXX^{1-145}$, in which missing side chains and loops were added, was created using Modeller (*24*), and scattering intensities of the model were simulated and compared with measured intensities using CRYSOL (*25*). Ten individual models of $PAXX^{1-145}$ created using DAMMIN (*26*) with the no-symmetry

operation were superimposed and averaged using SUPCOMB (*27*) and DAMAVER (*28*) (normalized spatial discrepancy: 0.566). The UCSF Chimera package (*29*) was used to visualize the averaged DAMMIN model and to fit crystallographic models into the envelopes.

Cells, siRNAs and transfections

293FT, U2OS and HeLa cells were cultured in Dulbecco's modified Eagle (DME) medium (Sigma-Aldrich) supplemented with 10% fetal bovine serum (BioSera), 2 mM L-glutamine, 100 units/ml penicillin and 100 μg/ml streptomycin (Sigma-Aldrich). RPE-1 cells were cultured in DME/F12 1:1 medium supplemented as above and additionally buffered with 0.2% $Na(CO_3)_2$. Stable U2OS and RPE-1 lines were established by selection in medium containing 500 μg/ml G418 (Life Technologies). RPE-1 cells expressing GFP or GFP-Ku70 from the endogenous locus were generated previously (*30*). Plasmids were transfected into cells using Lipofectamine 2000 (Life Technologies) or by electroporation using the Neon system according to the manufacturer's instructions (Life Technologies). siRNAs were transfected using Lipofectamine RNAiMAX (Life Technologies) according to the manufacturer's instructions, with the following sense sequences: siCtrl (targeting firefly luciferase), 5'-CGUACGCGGAAUACUUCGA-3'; siPAXX-1, 5'-CAGGAGAGUCGCUCAUCAA-3'; siPAXX-2, 5'-UAACACGGCUCCUCAAAUU-3'; siXRCC4, 5'-AUAUGUUGGUGAACUGAGA-3', siXLF 5'-CGCUGAUUCGAGAUCGAUUGA-3'.

Antibodies, SDS-PAGE and western blotting

Whole cell extracts were obtained by washing cells in phosphate-buffered saline (PBS) before lysis in SDS loading buffer (2% SDS, 10% (v/v) glycerol, 25 mM TCEP and 62.5 mM Tris-HCl, pH 6.8). Extracts were heated at 95 °C for 5 min, followed by shearing with 10 strokes through a 25G needle. Protein concentrations were determined by Bradford assay (Bio-Rad) or NanoDrop (Thermo Scientific). SDS-PAGE and western blotting were performed using the Novex NuPAGE SDS-PAGE gel system (Life Technologies) or the SE400 and TE42 systems from Hoefer. The following antibodies were used at the indicated dilutions: CHK1 (sc8408, Santa Cruz Biotechnology, 1/1000),

CHK1-pS345 (2348, Cell Signaling Technology, 1/10,000), CHK2-pT68 (2661, Cell Signaling Technology, 1/500), DNA-PKcs (MS-369-P1, Thermo Scientific, 1/200), GFP (11814460001, Roche, 1/5000), γH2AX (05-636, Millipore, 1/1000), H3 (ab1791, Abcam, 1/50,000), H3-pS10 (ab14955, Abcam, 1/5000), Ku70 (ab3114, Abcam, 1/1000), Ku80 (MS-285-P1, Thermo Scientific, 1/2000), p21 (sc397, Santa Cruz Biotechnology, 1/1000), p53 (554293, BD Biosciences, 1/6000), PAXX (ab126353, Abcam, 1/500), RPA1, (A300-241A, Bethyl Laboratories, 1/1000), TopBP1 (A300-111A, Bethyl Laboratories, 1/5000), XLF (ab33499, Abcam, 1/650) and XRCC4 (ab145, Abcam, 1/3000).

Immunofluorescence of fixed cells

For analyses of γH2AX foci, cells grown on glass coverslips were treated with 2 Gy of IR delivered by a CellRad system (Faxitron). Cells were fixed in 2% paraformaldehyde for 15 min at room temperature, permeabilized with PBS/0.2% Triton X-100 for 5 min and blocked with blocking buffer (5% BSA (w/v) and 0.1% Tween-20 in PBS). Cells were incubated with the following primary antibodies: γH2AX (05-636, Millipore, 1/1000) and Cyclin A (sc-751, Santa Cruz Biotechnology, 1/200), followed by secondary antibodies and 1 μg/ml DAPI in blocking buffer. Images were acquired using an Olympus FluoView 1000 confocal microscope. For analyses of centrosomal localization, HeLa cells grown on glass coverslips were transfected with pcDNA3.1(-) encoding Flag-PAXX and 24 hours later fixed with methanol at 4 ºC for 1 min. Cells were blocked with blocking buffer and incubated with the following primary antibodies: Flag (F7425, Sigma-Aldrich, 1/250) and γ-tubulin (T5326, Sigma-Aldrich, 1/500) followed by secondary antibodies and 0.5 μg/ml DAPI in blocking buffer. Images were acquired using an Applied Precision DeltaVision core microscope equipped with a Mercury 100 W lamp and CoolSnap HQ camera (Photometrics).

Peptide pulldown assays

Biotinylated $PAXX^{177\text{-}204}$ and $PAXX^{177\text{-}V199A/F201A\text{-}204}$ peptides were synthesised (Biomatik) and 250 μg was incubated with 400 μg of Dynabeads M-280 streptavidin (Life Technologies) equilibrated with 150 μl of TBS supplied with 0.1% (v/v) Igepal CA-

630 at 4 ºC for 1 hour. A peptide corresponding to the C-terminus of H2AX was used as a control (*31*). After washing with buffer six times, beads were suspended in 100 µl of buffer P1 (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 10% (v/v) glycerol, 5 mM NaF, 1 mM EDTA, 0.2% (v/v) Igepal CA-630, 1x cOmplete Protease Inhibitor Cocktail, EDTA free (Roche)). Beads were added to 500 µl of HeLa nuclear extract (Cilbiotech), which was diluted with the same volume of P1 and centrifuged at 17,200 x g, 4 ºC for 1 min. The mixture was incubated at 4 ºC for 3 hours and washed with P1 six times before 10 µl of 3x SDS-PAGE sample loading buffer was added. Pulldown samples were analysed by SDS-PAGE and the gel was visualized with silver stain (Pierce). Bands were analysed by GELC-MS/MS at the Cambridge Centre for Proteomics.

Immunoprecipitations

For preparation of lysates for immunoprecipitations (IPs), cells were washed twice in PBS, and lysed in IP buffer (100 mM NaCl, 0.2% (v/v) Igepal CA-630, 1 mM $MgCl_2$, 10% (v/v) glycerol, 5 mM NaF, 50 mM Tris-HCl, pH 7.5), supplemented with cOmplete EDTA-free protease inhibitor cocktail (Roche) and 25 U/ml Benzonase (Novagen). After nuclease digestion, NaCl and EDTA concentrations were adjusted to 300 mM and 2 mM, respectively, and lysates were cleared by centrifugation. Where appropriate, antibodies were added to a final concentration of 1 µg/mg lysate and incubated for 2 hours at 4 °C. Lysates were then incubated with 10 µl of GFP-Trap agarose beads (ChromoTek) or 20 µl of protein G Sepharose (GE Healthcare) for 2 hours with end-to-end mixing at 4 °C. Beads were washed five times in IP buffer before resuspension in 2x SDS loading buffer. 5% input lysate was loaded alongside unless otherwise stated.

Surface plasmon resonance

SPR experiments were performed using a Biacore T100 (GE Healthcare). 250 RU of 107 bp biotinylated DNA was immobilized using a wizard on channels 1 and 2 of an SA chip by following a recommended instruction (GE Healthcare). The DNA fragment was created using human *LIG4* gene as template by PCR reaction using a forward primer (5'-ATATATCCATGGCTGCCTCACAAACTTCACAA-3') and reverse primer (5'-biotin-TGCACGTCCTTTACTTTTCTGTAT-3'). 750 RU of Ku was then manually

immobilized on the channel 2 at 5 µl/min in 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM DTT, 0.005% (v/v) Tween-20. 5, 10, 20, 40 and 80 nM of $PAXX^{1-204}$ or $PAXX^{V199A/F201A}$ was flowed through channels 1 and 2 at 30 µl/min, 25 ºC for 60 s followed by washing with the buffer for 600 s using a kinetics wizard. A peptide-competition analysis was carried out in a similar way using 20 nM $PAXX^{1-204}$ with and without 20 nM-2 µM of $PAXX^{177-204}$ or 2 µM $PAXX^{177-V199A/F201A-204}$ (Biomatik). The dissociation constant of Ku and PAXX was calculated by BIAevaluation (GE Healthcare) assuming the Ku heterodimer and PAXX homodimer interact in a one-to one ratio (Fig. S9C).

Generation of $PAXX^{-/-}$ cells by CRISPR-Cas9

The PAXX knock-out strategy is displayed in Fig. S5A. Guide RNAs (gRNAs) targeting exon 4 of the *C9orf142* gene were selected using the Optimized CRISPR Design tool (http://crispr.mit.edu; (*32*)), with the following sequences: AGCCACAGCTTGCTGCTCAC (gRNA1) and AGCATCCCTGACGCTTTCAG (gRNA2). gRNA sequences were purchased as DNA oligonucleotides (Sigma-Aldrich) and cloned into pU6 gRNA cloning vector (a gift from Dr. W. Skarnes). The pMA-C9orf142 targeting vector (Fig. S5B) was constructed using the GeneArt plasmid construction service (Life Technologies). RPE-1 cells were transfected with pMA-C9orf142, human codon-optimized Cas9-D10A nickase (plasmid 41816, Addgene; (*33*)) and gRNA plasmids, and left to recover for 3 days before addition of selective medium containing 750 µg/ml G418. After 10 days, cells were selected based on GFP status using a MoFlo cell sorter (Beckman Coulter) and colonies were allowed to form from single cells. PAXX expression was analyzed by western blotting for 26 clones, of which 25 were found to have lost all detectable PAXX protein (data not shown). Genomic DNA (gDNA) from 2 cell clones was further analyzed by two rounds of PCR, firstly to confirm that the targeting cassette had integrated in the *C9orf142* locus and secondly, to decipher the fate of the second, non-targeted allele. gDNA was extracted and purified using a QIAamp DNA Mini Kit (Qiagen) according to the manufacturer's instructions. To confirm targeted integration of the cassette to replace the endogenous *C9orf142* gene, the following primers were used: TTCTTGCCGAGCCTTGAACA (forward, primer 1) and

GGACGACGGCAACTACAAGA (reverse, primer 2). To sequence the second allele, the following primers were used to amplify the region of the *C9orf142* gene containing the Cas9 cut site: GCGGCTTCAACCTCTAGTGA (forward, primer 3) and TGTGGGAGGGTACCTCTTGA (reverse, primer 4). Primer 4 was used to sequence the PCR product to characterize the disruption event at the second allele. Results from these analyses are shown in Fig. S5C.

Clonogenic survival assays

Where appropriate, cells were transfected with siRNAs and left to grow for 3 days. Cells were treated with ionized radiation delivered by a CellRad machine (Faxitron) or phleomycin (Melford Laboratories) for 2 hours, followed by 3 washes in PBS. Cells were left to form colonies for 7–14 days, and colonies were stained with 0.1% (w/v) crystal violet in 20% (v/v) ethanol for counting. Results were normalized to plating efficiencies of untreated cells.

Neutral comet assays

Where indicated, cells were transfected with siRNAs and left to grow for 3 days. Cells were mock-treated or damaged with 40 μg/ml phleomycin for 2 hours. For the recovery samples, cells were washed twice with PBS and cultured for a further 2 hours after phleomycin treatment. Cells were washed twice in $MgCl_2$/$CaCl_2$-free (-/-) PBS (Life Technologies) and harvested by scraping. Cells were resuspended at an approximate concentration of 5 x 10^6/ml in (-/-) PBS, and 10 μl of resuspended cells were mixed with 90 μl of LMAgarose (Trevigen) at 37 °C, before being spotted onto GelBond Film (Lonza), covered with a 22-mm cover slide (VWR International) and left at 4 °C for 10 min. Cover slides were removed, and cells were incubated in lysis solution (Trevigen) at 4 °C for 1 hour. Cells were washed in 90 mM Tris borate, pH 8.3 and 2 mM EDTA (TBE) and subjected to electrophoresis at 35 V for 7 min in TBE. Samples were fixed in 70% ethanol for 5 min and left to dry overnight. DNA was stained with SYBR Green I (Life Technologies) in 10 mM Tris-HCl (pH 7.5) and 1 mM EDTA for 10 min, and left to dry overnight. Images were taken with an IX71 fluorescent microscope and Cell^F software (Olympus). Tail moments were measured using CometScore software (TriTek).

Mean tail moments were measured and calculated from at least 50 cells per condition. Efficiency of DSB repair was determined as the mean tail moment ratio between DNA from recovered and damaged cells.

Random plasmid integration NHEJ assays

Two days after siRNA transfection, cells were transfected with *Bam*HI/*Xho*I-linearized pEGFP-C1, before being plated at high and low densities the next day. Cells at high densities were treated with 1 mg/ml G418 1 day after plating, and colonies were allowed to grow for 10-14 days. Cells were stained with 0.1% (w/v) crystal violet in 20% ethanol for counting. Random plasmid integration events were normalized for plating efficiency and normalized to siCtrl samples to obtain final values for NHEJ efficiency.

Laser micro-irradiation and live cell imaging

Assays were performed as described previously (*34*).

Electrophoretic mobility shift assays (EMSAs)

50-bp dsDNA labelled with 6-FAM(6-carboxyfluorescein) was created by annealing a forward primer (5'-6FAM-TAAATGCCAATGCTGCTGATACGTACTCGGACTGATTCGGAACTGTAACG -3') and reverse primer (5'-CGTTACAGTTCCGAATCAGTCCGAGTACGTATCAGCAGCATTGGCATTTA-3'). For the assays using the NHEJ ligase complex, 10 nM of 50-bp DNA was incubated with proteins as indicated in the main text in 20 mM Tris-HCl pH 7.5, 50 mM KCl, 5% (v/v) glycerol, 100 μM DTT, 10 μg/ml BSA. Samples were incubated at room temperature for 30 min and applied onto a 5% polyacrylamide gel in 0.5x TBE buffer for electrophoresis. DNA was visualized by Typhoon 9000 (GE Healthcare). The DNA-binding assay using pFastBacHTB digested with *Hind*III and *Bam*HI was carried in a similar manner. The incubated samples were applied to a 0.8% (w/v) agarose gel in 0.5x TBE buffer for electrophoresis. DNA was stained with SYBR Gold (Life Technologies) and visualized using a UV imager.

DNA ligation assays

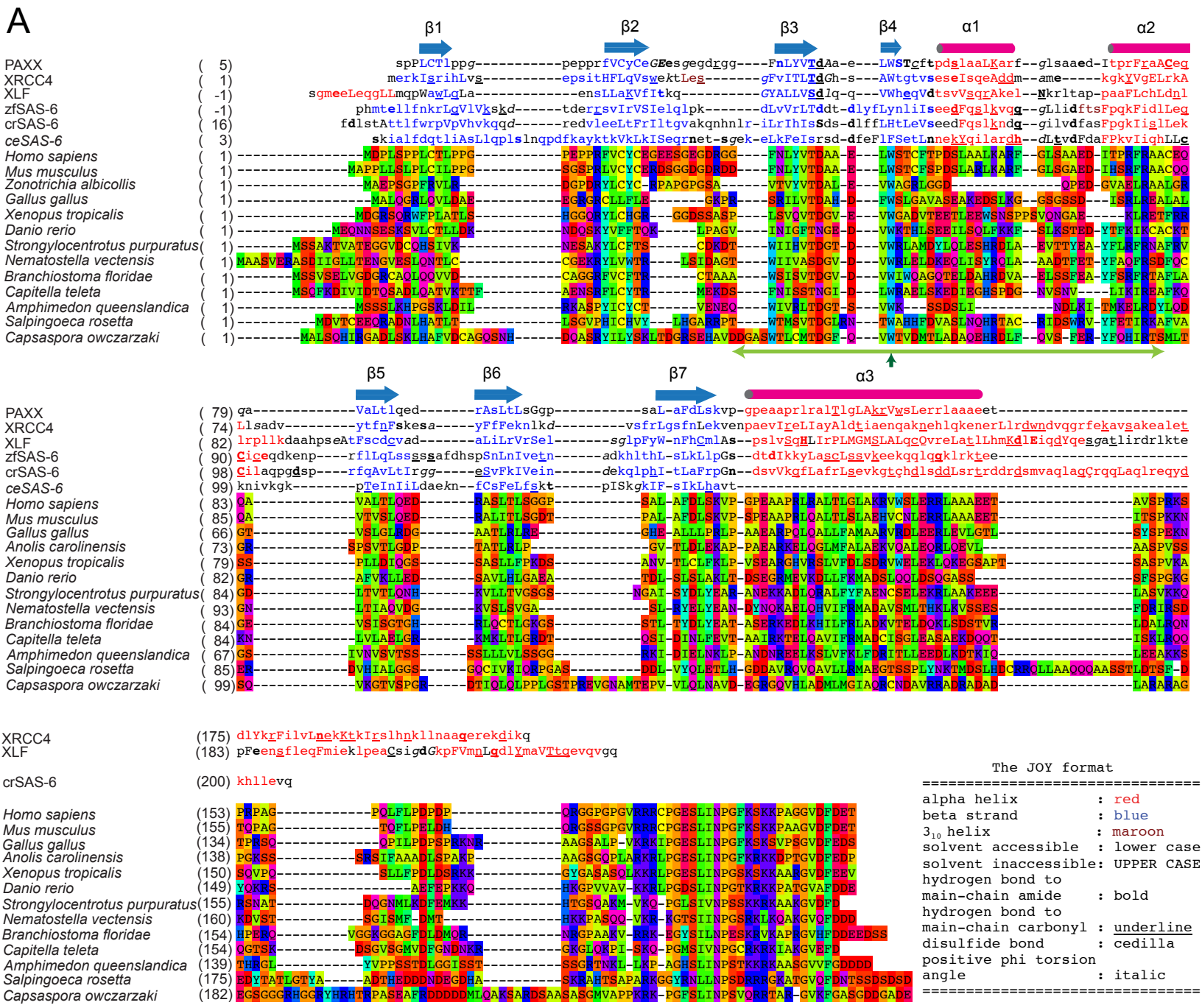
pcDNA3.1(-) was digested with *Eco*RV (New England BioLabs) and purified using NucleoSpin Gel and PCR Clean-up (Clontech). 50 ng of the digested plasmid was mixed with the indicated quantities of proteins in 20 µl of reaction buffer containing 25 mM Tris-HCl pH 7.5, 150 mM KCl, 1 mM $MgCl_2$, 1 mM DTT, 10 µM ATP, 10% (w/v) PEG10,000, 10 µg/ml BSA. Mixtures were incubated at 37 ºC for 5 min before initiating ligation by XRCC4/LIG4 at 37 ºC for 30 min. The mixtures were incubated at 50 ºC for another 30 min after adding 2 µl of a reaction-stop solution (100 mM EDTA, 0.1% (w/v) SDS) and 0.2 µl of 20 mg/ml Proteinase K. Reaction mixtures were applied onto 0.8% agarose gel in TBE buffer for electrophoresis. The gel was stained with SYBR Gold, visualized using a UV imager and quantified using GeneTools (Synegene).

Chromatin fractionation

Cells were mock-treated or treated for 1 hour with 300 µM phleomycin to induce large numbers of DSBs, before being harvested either at this point or after an additional 1-hour recovery period in phleomycin-free medium after three PBS washes. Cells were washed twice in ice-cold PBS, and pre-extracted twice for 3 min at 4 °C in CSK buffer (100 mM NaCl, 300 mM sucrose, 3 mM $MgCl_2$, 0.7% Triton X-100, 10 mM PIPES, pH 7.0), supplemented with 0.3 mg/ml RNase A (Sigma-Aldrich). After pre-extraction, cells were washed 3 times in ice-cold PBS and harvested in SDS loading buffer for analysis by SDS-PAGE and western blotting.

Figure S1

A



The JOY format

```
==================================
alpha helix          : red
beta strand          : blue
3_10 helix           : maroon
solvent accessible   : lower case
solvent inaccessible: UPPER CASE
hydrogen bond to
main-chain amide     : bold
hydrogen bond to
main-chain carbonyl  : underline
disulfide bond       : cedilla
positive phi torsion
angle                : italic
==================================
```

B

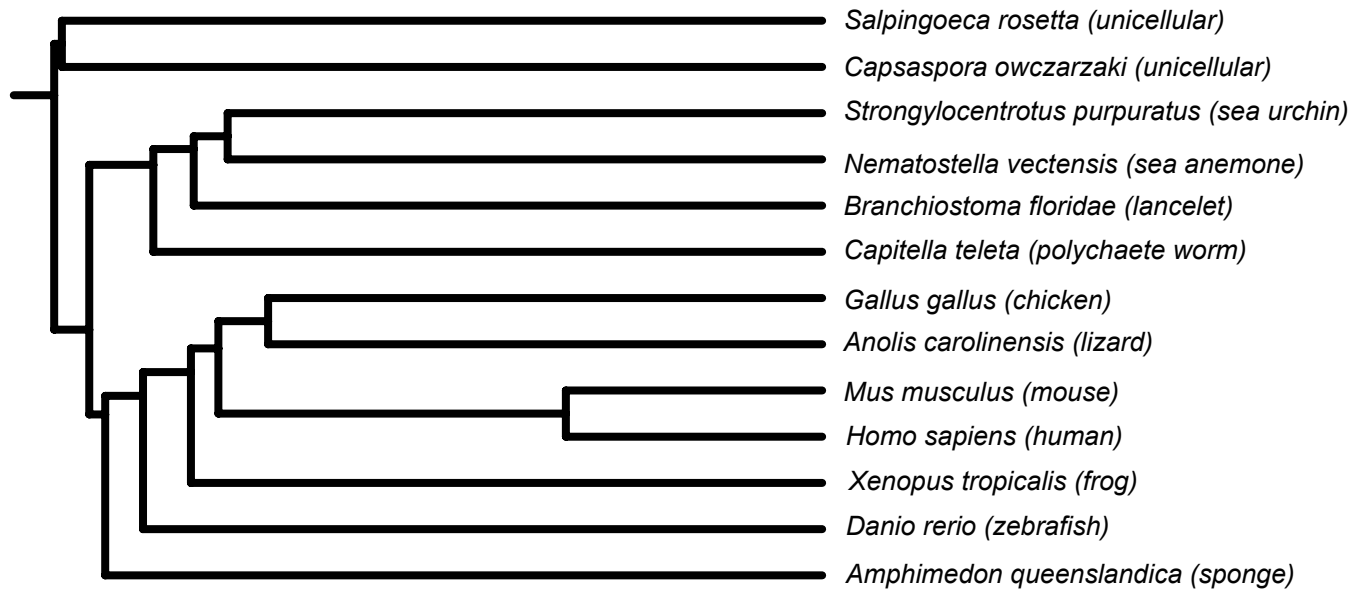


Fig. S1. Evolutionary analysis of PAXX.

(**A**) Sequence alignment of XRCC4, XLF, SAS6 and PAXX orthologs. Structures of XRCC4, XLF and SAS6 (PDB codes: 1IK9 (*7*), 2QM4 (*9*), 2Y3W (zfSAS6) (*10*), 3PYI (ceSAS6) and 3Q0X (crSAS6) (*11*)) were aligned together with PAXX orthologs from various organisms, as indicated. Since the sequence of a PAXX homologue of *Capsaspora* was fused with the C-terminus of a predicted Co/Zn/Cd cation transporter in the NCBI database (NCBI Reference Sequence: XP_004364434.1), we concluded that M381 of the chimeric protein was the first methionine of the PAXX ortholog. The structural alignment is in JoY format (*35*) as indicated at the end of the alignment and the sequence alignment is represented in the Taylor color scheme. Secondary structure elements of the PAXX structure are shown above the alignment. PISA motifs and conserved tryptophans (see main text) are indicated by horizontal green and vertical black arrows, respectively. (**B**) A phylogenetic tree of PAXX orthologs,generated using TraceSuiteII (*36*). By using PSI-BLAST (*37*) followed by FUGUE searches (*38*), we identified PAXX homologs only in Filozoa, which includes animals and unicellular organisms such as choanoflagellates and the monotypic genus *Capsaspora*.

Figure S2

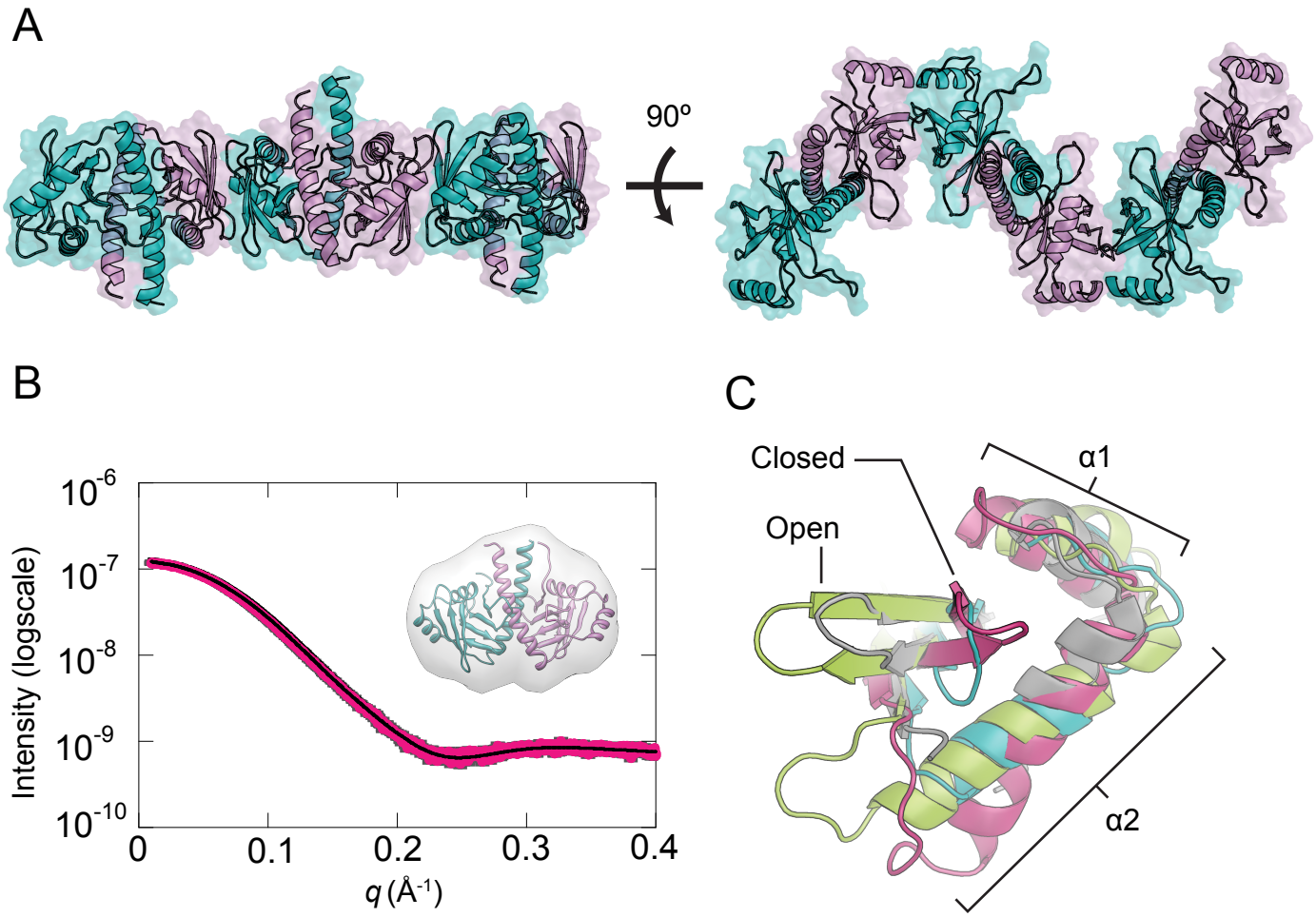

A
90°
B
Intensity (logscale)
10-6
10-7
10-8
10-9
10-10
0
0.1
0.2
0.3
0.4
q (Å-1)
C
Closed
Open
α1
α2

Fig. S2. PAXX is a dimer in solution.

(A) Zigzag filament structure of PAXX. The coiled-coil of the PAXX homodimer (cyan and pink) is normal to and pointing out of the page in the left and right projections, whereas those of the other two dimers point in the opposite direction. **(B)** SAXS studies of $PAXX^{1-145}$. Pink dots with error bars (grey) represent measured intensity and the black line in the scattering profile is calculated intensity from the crystal structure of $PAXX^{1-145}$ (χ=1.166). Inset shows an averaged DAMMIN model (grey) of $PAXX^{1-145}$ together with the crystal structure. **(C)** Comparison of XRCC4-superfamily members focusing on closed and open conformations of the head-to-head interaction regions. PAXX, XRCC4, XLF and SAS6 are shown in cyan, silver, magenta and lime respectively.

Figure S3

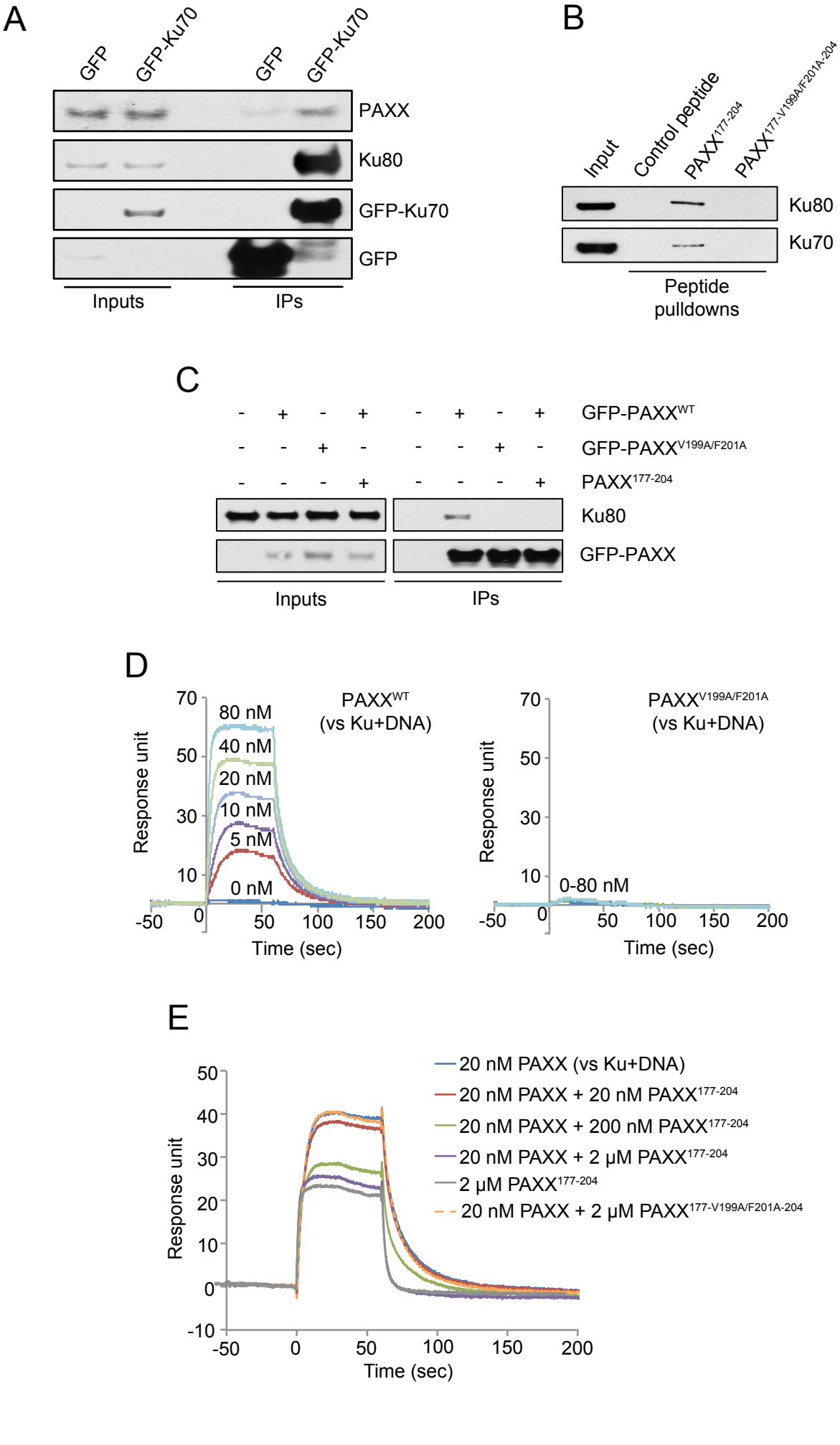
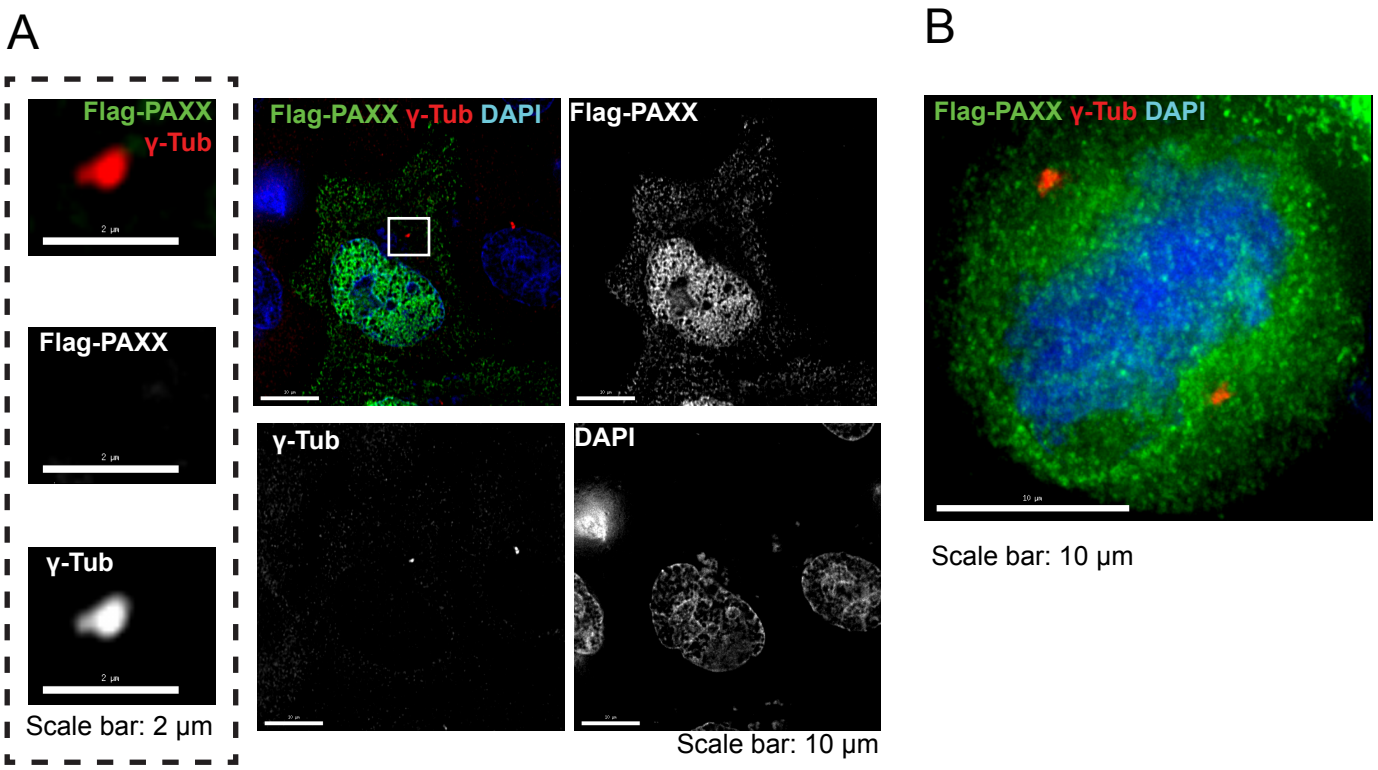
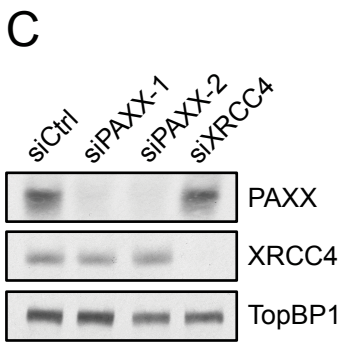

A
GFP
GFP-Ku70
GFP
GFP-Ku70
PAXX
Ku80
GFP-Ku70
GFP
Inputs
IPs
B
Input
Control peptide
PAXX177-204
PAXX177-V199A/F201A-204
Ku80
Ku70
Peptide pulldowns
C
GFP-PAXXWT
GFP-PAXXV199A/F201A
PAXX177-204
Ku80
GFP-PAXX
Inputs
IPs
D
PAXXWT (vs Ku+DNA)
80 nM
40 nM
20 nM
10 nM
5 nM
0 nM
Response unit
Time (sec)
PAXXV199A/F201A (vs Ku+DNA)
0-80 nM
Response unit
Time (sec)
E
20 nM PAXX (vs Ku+DNA)
20 nM PAXX + 20 nM PAXX177-204
20 nM PAXX + 200 nM PAXX177-204
20 nM PAXX + 2 μM PAXX177-204
2 μM PAXX177-204
20 nM PAXX + 2 μM PAXX177-V199A/F201A-204
Response unit
Time (sec)

Fig. S3. PAXX interacts directly with Ku.

(**A**) Reciprocal GFP-pulldown from RPE-1 cells expressing GFP-tagged Ku70 from the endogenous *XRCC6* locus showing that Ku70 interacts with PAXX. **(B)** WT C-terminal PAXX peptide, but not the V199A/F201A mutant, interacts with Ku. (**C**) Addition of $PAXX^{177-204}$ peptide to cell lysate disrupts Ku-PAXX complexes in U2OS cells stably expressing the indicated GFP-PAXX protein where appropriate. **(D)** SPR of Ku-DNA and PAXX, showing that they interact with a K_d of 14.5 nM (assuming a 1:1 interaction between Ku-DNA and a PAXX homodimer). The direct interaction between Ku-DNA and $PAXX^{1-204}$ was measured by comparing two channels immobilised Ku-DNA and DNA only in the presence of 0-80 nM of $PAXX^{WT}$ (left) or $PAXX^{V199A/F201A}$ (right). (**E**) SPR profiles of $PAXX^{177-204}$ peptide competing with full-length recombinant PAXX for the Ku-DNA interaction. $PAXX^{1-204}$ was added to immobilized Ku on DNA in the presence and absence of $PAXX^{177-204}$ or $PAXX^{177\text{-}V199A/F201A\text{-}204}$.

Figure S4


A
Flag-PAXX
γ-Tub
Flag-PAXX
γ-Tub
Scale bar: 2 μm
Flag-PAXX γ-Tub DAPI
Flag-PAXX
γ-Tub
DAPI
Scale bar: 10 μm
B
Flag-PAXX γ-Tub DAPI
Scale bar: 10 μm


C
siCtrl
siPAXX-1
siPAXX-2
siXRCC4
PAXX
XRCC4
TopBP1

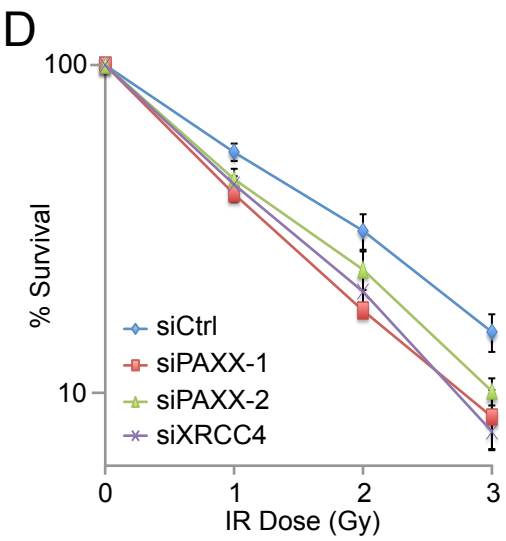
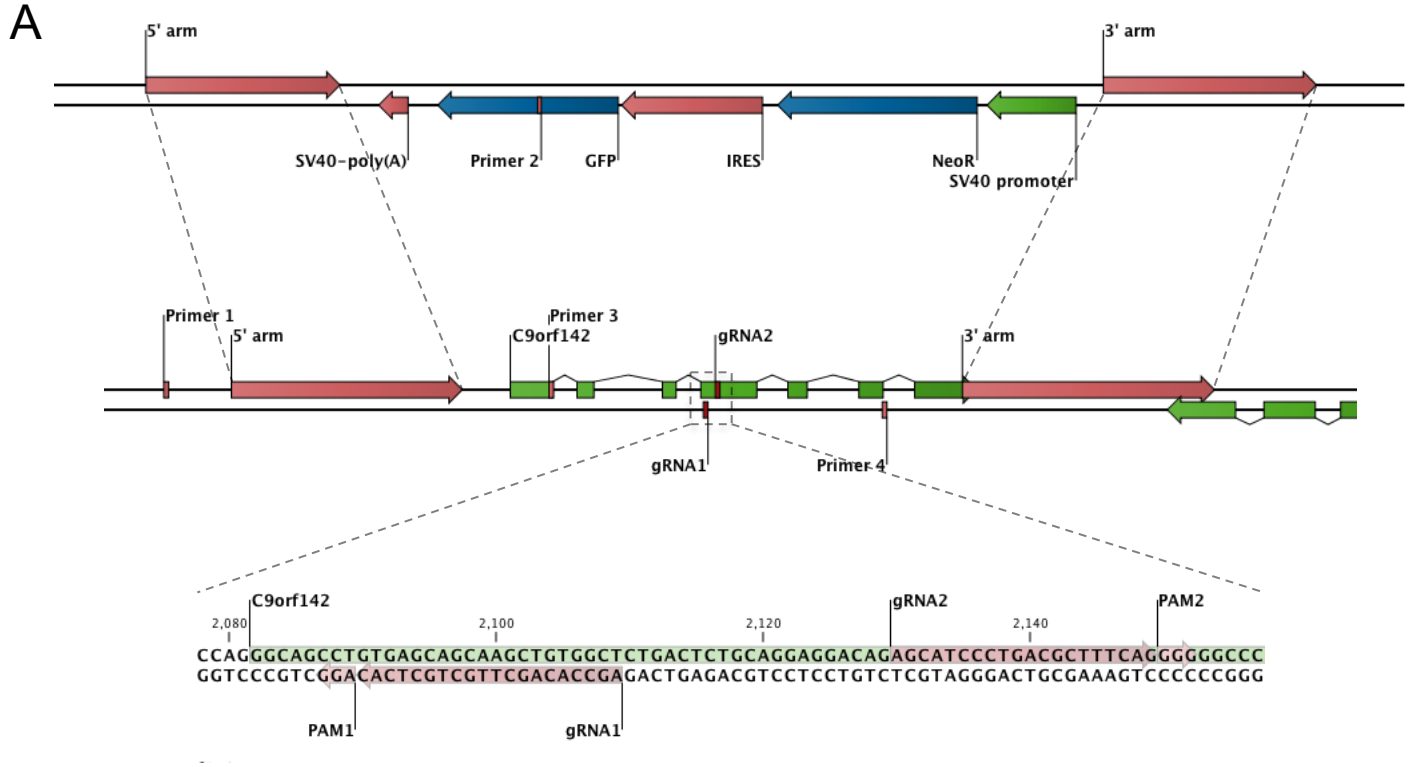

D
100
10
% Survival
siCtrl
siPAXX-1
siPAXX-2
siXRCC4
0
1
2
3
IR Dose (Gy)

Fig. S4. PAXX localizes in the nucleus and PAXX-deficient cells are hypersensitive to IR.

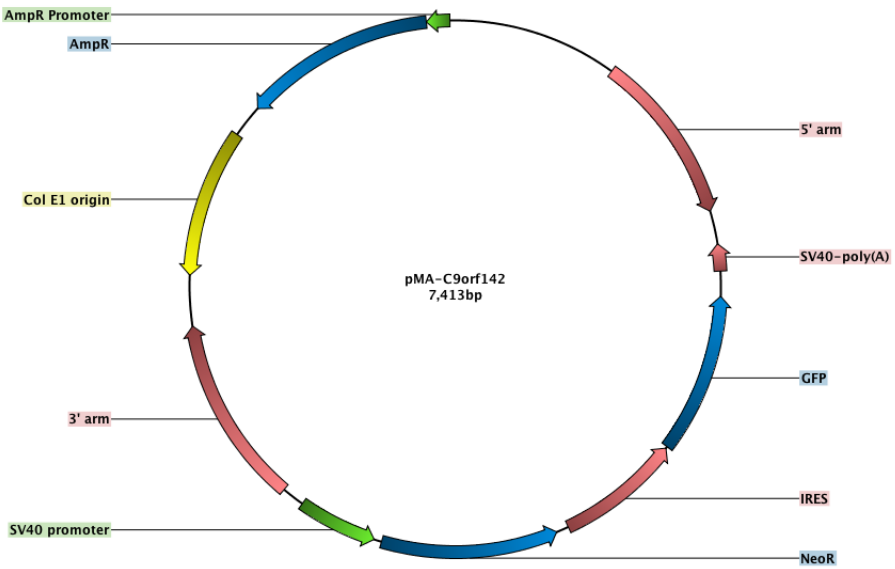
(**A, B**) PAXX is predominantly nuclear and does not localise to interphase or mitotic centrosomes. Interphase (A) and mitotic (B) images of HeLa cells transfected with plasmid encoding Flag-PAXX and immunostained with antibodies against Flag, γ-tubulin (centrosome marker) and DAPI (DNA dye). Magnified images in A (left) are single-plane images of the area marked in white box. Projected image of 41 z-planes is shown in B. (**C**) Representative western blot showing efficiency of PAXX and XRCC4 depletion by siRNAs in the U2OS cells used in subsequent experiments. TopBP1 is a loading control. (**D**) PAXX depletion causes radiosensitivity similar to that caused by XRCC4 depletion in clonogenic survival assays. Error bars represent the SEM from 3 independent replicates in this and subsequent experiments.

Figure S5

A



B



C

Clone	Indel	Indel sequence	Size (nt)	Predicted consequences for PAXX protein
1	del397-690	CTGTAGCTGTAGTCCTCCCATTGGCTAGGGCTCTGGGGTCGGGC AGGTTTCGGGTGCCCCCAGTGGGCCTCGGGTTCCAGGCAGCTC GTGACAAGCCCCTGTGCTCTCTAGAAAGCCCGTTTTGGCCTGAG TGCGGCTGAGGACATCACCCCCCGGTTCAGGTGAGACCCAAGC AGGGAGGAAGGACGGGTGGGGAGGAGGAGGGTCTGCCACAGC TCTCCGCACCTCTCCTCTCCCAGGGCAGCCTGTGAGCAGCAAGC TGTGGCTCTGACTCTGCAGGAGGACAGAGCATCC	294	Deletion causes frameshift and loss of PAXX residues after Leu60, assuming that the disrupted exon 4 is spliced out
2	in679	TGGCTAGCTTGACTCTGCAGGAGTGGCTCTGTCCTCTGCAGGAG	44	Insertion causes frameshift after Glu91 and premature stop codon after 3 additional amino acids

Fig. S5. Generation of PAXX$^{-/-}$ RPE-1 cells using CRISPR-Cas9.

(**A**) Schematic representation of the *C9orf142* gene targeting strategy. (**B**) Schematic representation of the *C9orf142* gene targeting vector pMA-C9orf142. (**C**) Table summarizing the disruption of the second allele in 2 independent PAXX$^{-/-}$ cell clones.

Figure S6

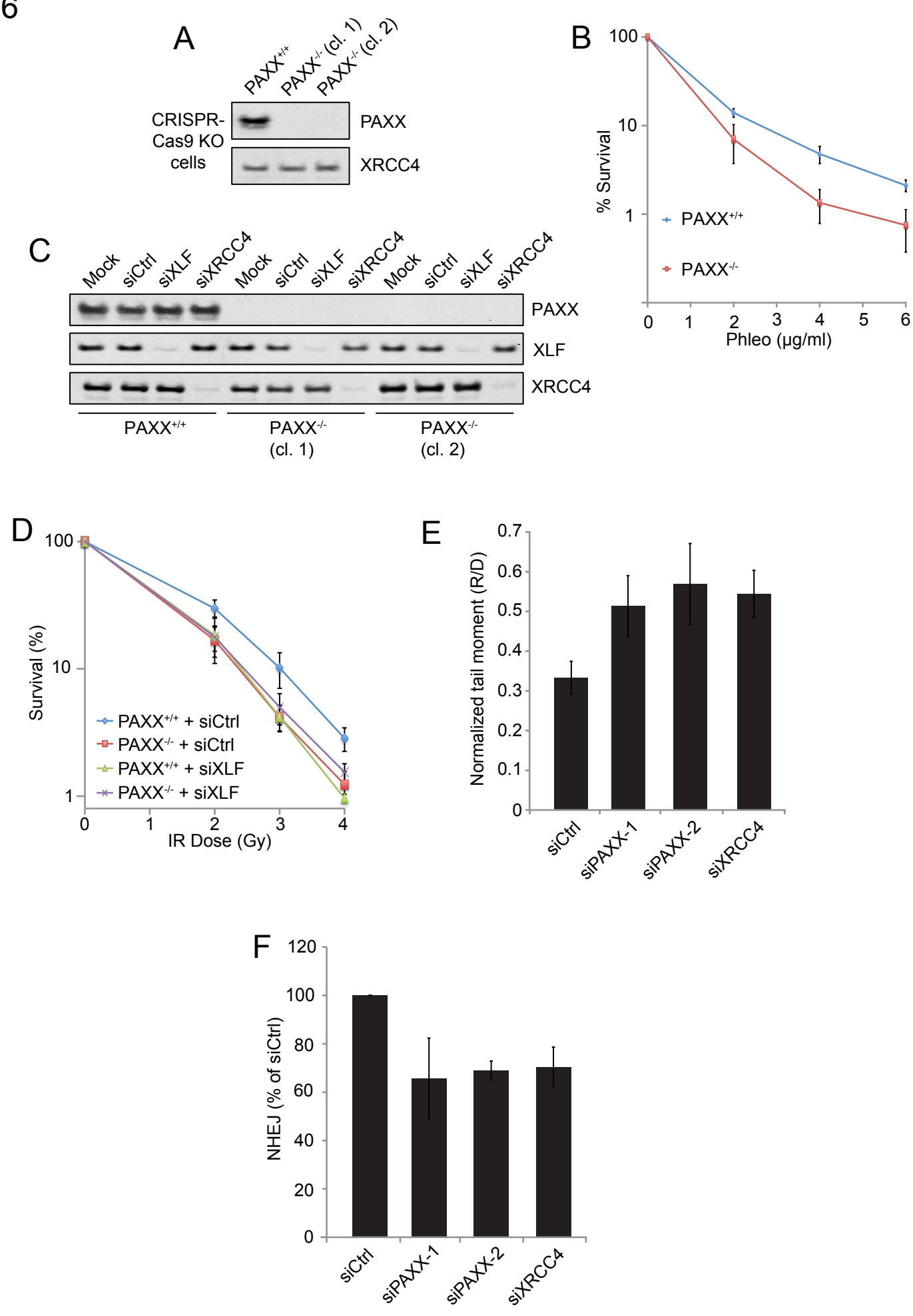

A
PAXX+/+
PAXX-/- (cl. 1)
PAXX-/- (cl. 2)
CRISPR-Cas9 KO cells
PAXX
XRCC4
B
% Survival
100
10
1
PAXX+/+
PAXX-/-
0 1 2 3 4 5 6
Phleo (µg/ml)
C
Mock siCtrl siXLF siXRCC4 Mock siCtrl siXLF siXRCC4 Mock siCtrl siXLF siXRCC4
PAXX
XLF
XRCC4
PAXX+/+
PAXX-/- (cl. 1)
PAXX-/- (cl. 2)
D
Survival (%)
100
10
1
PAXX+/+ + siCtrl
PAXX-/- + siCtrl
PAXX+/+ + siXLF
PAXX-/- + siXLF
0 1 2 3 4
IR Dose (Gy)
E
Normalized tail moment (R/D)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
siCtrl
siPAXX-1
siPAXX-2
siXRCC4
F
NHEJ (% of siCtrl)
120
100
80
60
40
20
0
siCtrl
siPAXX-1
siPAXX-2
siXRCC4

Fig. S6. PAXX-deficient cells are hypersensitive to DSB-inducing agents and are defective in DSB repair.

(**A**) Western blot confirming that PAXX expression is absent in PAXX$^{-/-}$ RPE-1 cells. XRCC4 is a loading control. In this and subsequent figures, "cl." indicates clone number. (**B**) PAXX$^{-/-}$ RPE-1 cells are hypersensitive to phleomycin in clonogenic survival assays. (**C**) Representative western blot showing efficiency of PAXX, XLF and XRCC4 depletion by siRNAs in PAXX$^{+/+}$ and PAXX$^{-/-}$ RPE-1 cells used in subsequent experiments. (**D**) Clonogenic survival assay showing that PAXX$^{-/-}$ cells are radiosensitive and that PAXX loss is epistatic with XLF depletion. Note that this is the same experiment as shown in Fig. 3D of the main manuscript as both experiments were carried out concurrently. **(E)** Neutral comet assay showing that PAXX depletion in U2OS cells causes a similar DSB repair defect as that observed in XRCC4-depleted cells or PAXX$^{-/-}$ RPE-1 cells. **(F)** PAXX is required for NHEJ as measured by random plasmid integration in U2OS cells transfected with the indicated siRNAs.

Figure S7

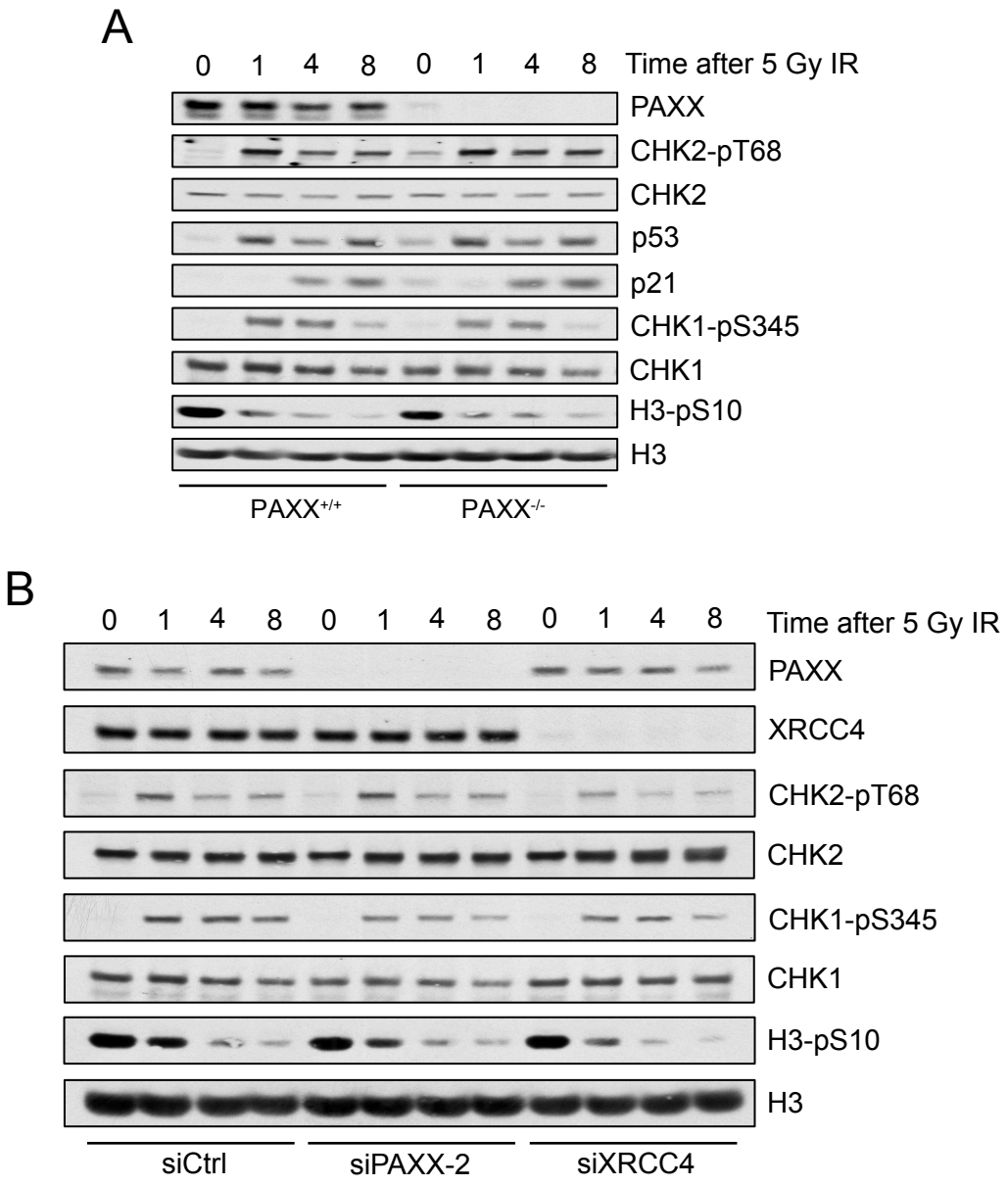

A
0 1 4 8 0 1 4 8 Time after 5 Gy IR
PAXX
CHK2-pT68
CHK2
p53
p21
CHK1-pS345
CHK1
H3-pS10
H3
PAXX+/+
PAXX-/-
B
0 1 4 8 0 1 4 8 0 1 4 8 Time after 5 Gy IR
PAXX
XRCC4
CHK2-pT68
CHK2
CHK1-pS345
CHK1
H3-pS10
H3
siCtrl
siPAXX-2
siXRCC4

Fig. S7. PAXX is not required for checkpoint signaling.

(**A**) Western blots showing that induction of checkpoint signaling is similar in $PAXX^{+/+}$ and $PAXX^{-/-}$ RPE-1 cells. (**B**) Western blots from U2OS cell extracts showing that checkpoint signaling is similar to control cells regardless of PAXX or XRCC4 status.

Figure S8

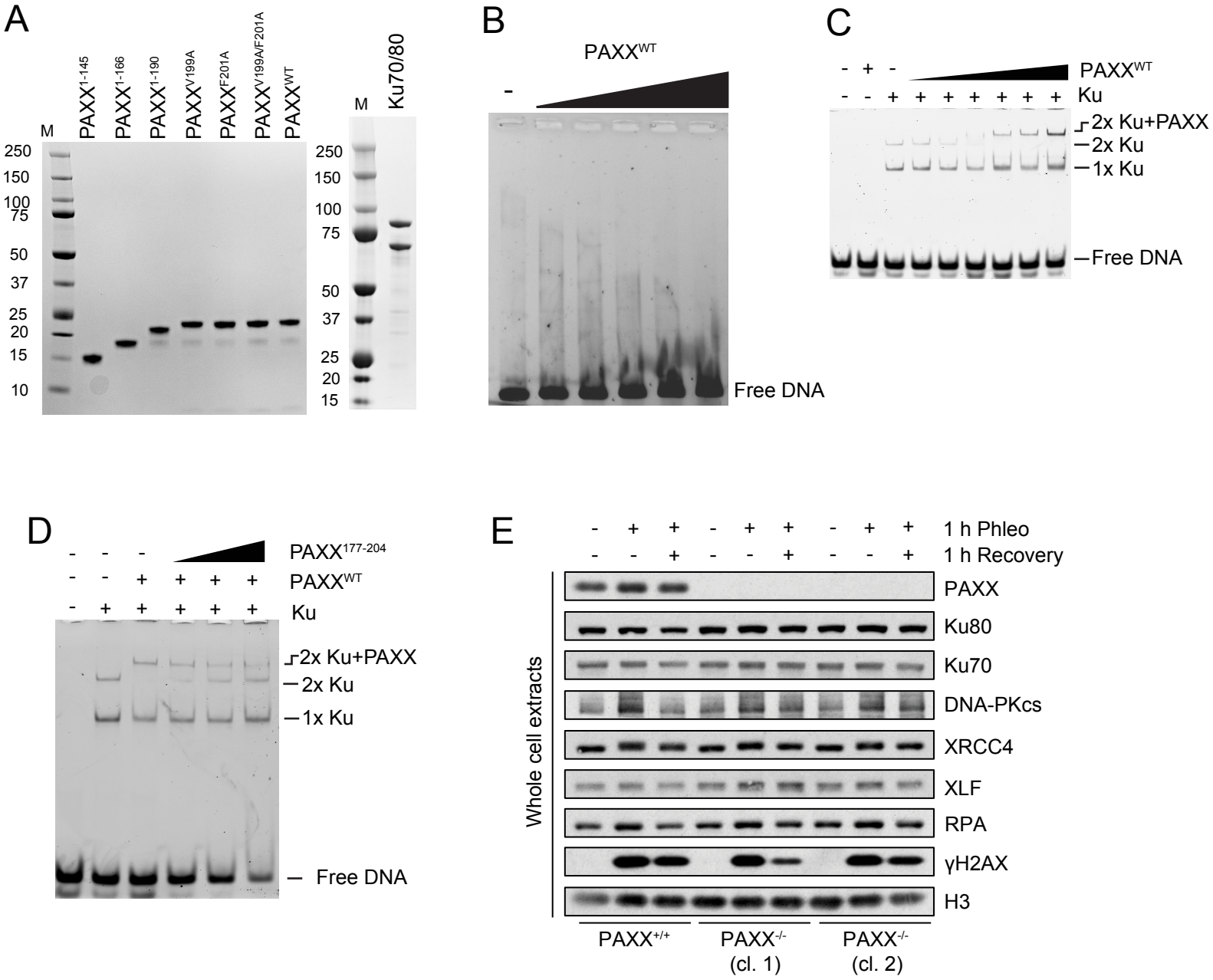
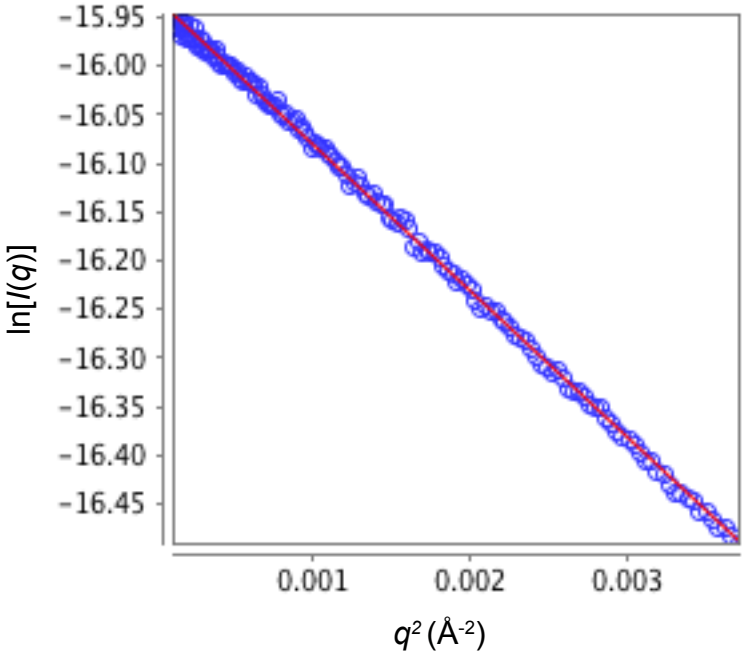

A
M
PAXX1-145
PAXX1-166
PAXX1-190
PAXXV199A
PAXXF201A
PAXXV199A/F201A
PAXXWT
250
150
100
75
50
37
25
20
15
10
M
Ku70/80
B
PAXXWT
-
Free DNA
C
PAXXWT
Ku
2x Ku+PAXX
2x Ku
1x Ku
Free DNA
D
PAXX177-204
PAXXWT
Ku
2x Ku+PAXX
2x Ku
1x Ku
Free DNA
E
1 h Phleo
1 h Recovery
PAXX
Ku80
Ku70
DNA-PKcs
XRCC4
XLF
RPA
γH2AX
H3
Whole cell extracts
PAXX+/+
PAXX-/-
(cl. 1)
PAXX-/-
(cl. 2)

Fig. S8. PAXX interacts with Ku *in vitro* and does not affect overall levels of NHEJ factors.

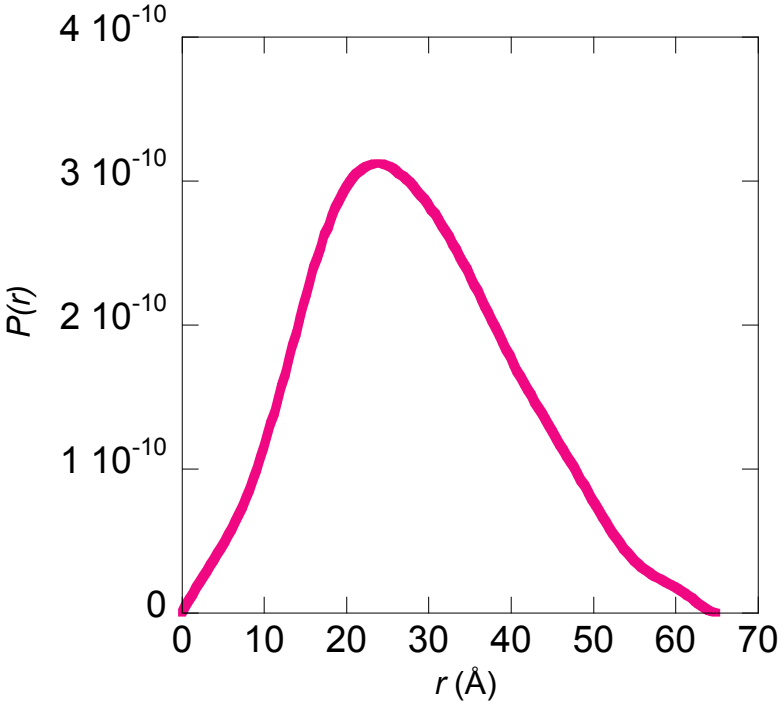
(**A**) Purified PAXX and Ku proteins. 2 μg of recombinant protein was loaded on each lane. "M" indicates protein ladder (kDa). (**B**) DNA-binding assays with recombinant PAXX. 100 ng of linearized pFastBacHTB plasmid was incubated with 1, 2, 3, 4 or 5 μg of $PAXX^{1-204}$. Agarose gel was stained with SYBR gold. (**C**) EMSA of Ku and $PAXX^{1-204}$. 20 nM of Ku was added where indicated. The concentration of $PAXX^{1-204}$ was increased sequentially 2-fold from 20 nM. For EMSA without Ku, 640 nM of $PAXX^{1-204}$ was added. (**D**) EMSA of Ku and $PAXX^{1-204}$ with PAXX peptide. 20 nM of Ku and 200 nM of $PAXX^{1-204}$ were added where indicated. The concentration of $PAXX^{177-204}$ was increased by 5-fold from 4 μM. (**E**) Whole cell extract controls for Fig. 4C in the main manuscript.

Figure S9

A



B



C

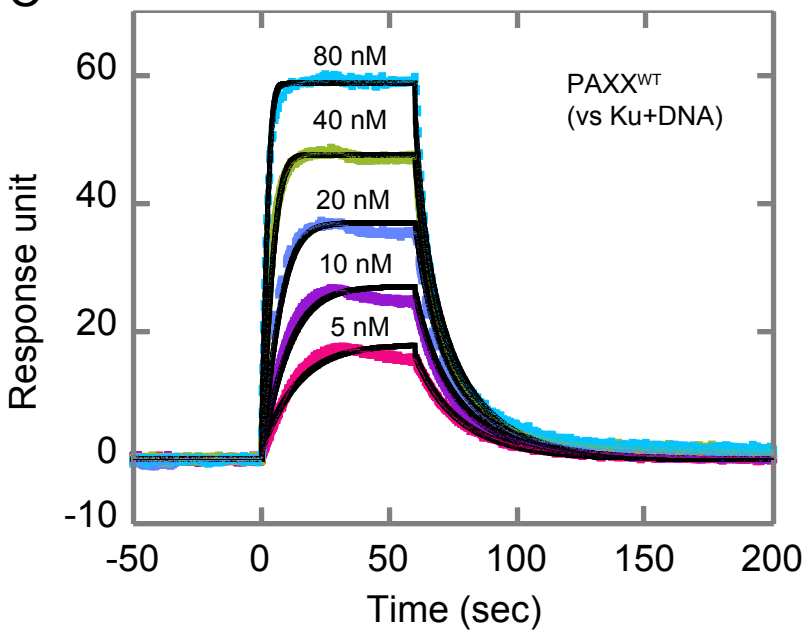


Fig. S9. Biophysical characterization of PAXX.

(**A**) Monodispersity of $PAXX^{1-145}$ in solution. Guinier plot of $PAXX^{1-145}$ generated using ScÅtter shows the linearity (pink solid line) of observed intensity (blue circles) of the Guinier region of $PAXX^{1-145}$. (**B**) The *P*(*r*) distribution function of $PAXX^{1-145}$. (**C**) Fitting of SPR profiles of $PAXX^{1-204}$ bound on Ku. Fitting curves (black) are overlaid over the SPR profiles of $PAXX^{1-204}$ shown in Fig. S3D.

Table S1.

Result of Back-Phyre search. Back-Phyre search (*39*) was carried out using the head domain of human XRCC4 (residues 1-118; PDB code: 1IK9). C9orf142 is highlighted in yellow.

Rank	Description	Alignment Coverage	Confidence	% i.d.
1	gi|4507945|ref|NP_003392.1| X-ray repair cross complementing protein 4 isoform 1 [Homo sapiens]	1-114	100	99
2	gi|12408649|ref|NP_072044.1| X-ray repair cross complementing protein gi|12408649|ref|NP_072044.1| X-ray repair cross complementing protein	1-114	100	99
3	gi|12408647|ref|NP_071801.1| X-ray repair cross complementing protein 4 isoform 2 [Homo sapiens]	1-114	100	99
4	gi|13376142|ref|NP_079058.1| XRCC4-like factor [Homo sapiens]	30-103	80.5	20
5	gi|29789058|ref|NP_056053.1| KIBRA protein [Homo sapiens]	13-39	54.6	15
6	gi|7662404|ref|NP_055774.1| caspase recruitment domain family, member 8 [Homo sapiens]	14-47	36	15
7	gi|48255889|ref|NP_002734.2| protein kinase C substrate 80K-H isoform 1 [Homo sapiens]	3-45	35.6	16
8	gi|39930541|ref|NP_899064.1| hypothetical protein LOC286257 [Homo sapiens]	27-107	34.1	11
9	gi|4507737|ref|NP_003314.1| tubby like protein 2 [Homo sapiens]	17-111	29.8	18
10	gi|48255891|ref|NP_001001329.1| protein kinase C substrate 80K-H isoform 2 [Homo sapiens]	3-45	28.6	16
11	gi|4505525|ref|NP_002544.1| origin recognition complex subunit 5 isoform 1 [Homo sapiens]	43-69	28	26
12	gi|14249738|ref|NP_115909.1| N-acetylglucosamine-1-phosphotransferase, gamma subunit [Homo sapiens]	16497	25.1	12
13	gi|19263349|ref|NP_597724.1| Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 4 [Homo sapiens]	47-82	20.9	15

Table S2.

Result of FUGUE search of C9orf142. Structures predicted to be certain homologues are highlighted in yellow.

PDB code (chain) or protein family	Z-score	Certainty
1IK9 (A)	8.28	CERTAIN
1FU1 (A)	6.7	CERTAIN
2QM4 (A)	4.67	LIKELY
3AEI (A)	4.2	LIKELY
1ZXA (A)	3.99	MARGINAL
1WNL (A)	3.89	MARGINAL
Fibrinogen C	3.3	GUESS
2KRG (A)	2.9	GUESS
3A2A (A)	2.85	GUESS
2OSZ (A)	2.74	GUESS

Table S3.

Statistics of X-ray diffraction data from crystals and structural refinement of PAXX constructs.

Crystal	$PAXX^{1-145}$	$PAXX^{1-166}$	$PAXX^{1-204}$
	Derivative $KAu(CN)_2$ (Peak)	Native	Native
X-ray Source	Diamond Light Source, Beamline I03		
Wavelength (Å)	1.03961	0.97625	0.97625
Resolution (Å)	45.52-2.46	55.20-2.35	24.47-3.45
Space group	$P6_522$		
Unit Cell Parameters			
$a=b$ (Å)	91.04	91.95	89.58
c (Å)	152.7	153.2	151.4
β (°)	120	120	120
Number of unique reflections	14,219	16,230	4,986
Completeness (%)	100 (100)[a]	98.5 (99.3)	97.3 (98.2)
Redundancy	18.3	9.6	6.1
[b]R_{sym} (%)	4.0 (76.5)	5.2 (62.7)	8.8 (59.9)
<I/σ>	40.4 (4.4)	24.3 (4.4)	9.1 (1.7)
Phasing			
F.O.M.	0.311		
Overall score	49.85		

Refinement			
PDB code		3WTD	3WTF
R_{cryst}[c] (highest shell)		19.4 (23.0)	23.5 (36.6)
R_{free}[d] (highest shell)		22.9 (26.1)	26.8 (41.8)
Number of atoms:			
Protein atoms		1970	1809
Water molecules		97	3
Average B-factors (Å^2)		67.3	168.8
Ramachandran favored (%)		97.7	97.6
R.M.S.D.			
Bond (Å)		0.008	0.012
Angle (˚)		1.131	1.642

[a]The statistics in parenthesis are for the highest resolution shell

[b]$R_{sym}=\Sigma_h|I_h - \langle I \rangle|/\Sigma_h I_h$, where I_h is the intensity of reflection h, and $\langle I \rangle$ is the mean intensity of all symmetry-related reflections.

[c]$R_{cryst}=\Sigma||F_{obs}|-|F_{calc}||/\Sigma|F_{obs}|$, F_{obs} and F_{calc} are observed and calculated structure factor amplitudes.

[d]R_{free} as for R_{cryst} using a random subset of the data (about 5% for PAXX$^{1-145}$ and 10% for PAXX$^{1-204}$) excluded from the refinement.

Table S4.
ES-MS and SAXS studies of PAXX.

Protein	R_g (Å)	D_{max} (Å)	Measured Mass by ES-MS (Da)	Theoretical MW of monomer (Da)
$PAXX^{1-145}$	21.66	65	31216.4 ±0.6	15608.6
$PAXX^{1-204}$	N/A	N/A	43566.4 ±1.7	21783.6

References

14. J. Peränen, M. Rikkonen, M. Hyvönen, L. Kääriäinen, T7 Vectors with a Modified T7lacPromoter for Expression of Proteins inEscherichia coli, *Anal. Biochem.* **236**, 371–373 (1996).

15. Y. Wang, B. J. Lamarche, M.-D. Tsai, Human DNA Ligase IV and the Ligase IV/XRCC4 Complex: Analysis of Nick Ligation Fidelity, *Biochemistry* **46**, 4962–4976 (2007).

16. L. A. Hanakahi, 2-Step purification of the Ku DNA repair protein expressed in Escherichia coli, *Protein Expr. Purif.* **52**, 139–145 (2007).

17. W. Kabsch, Integration, scaling, space-group assignment and post-refinement., *Acta Crystallographica. Section D, Biological crystallography* **66**, 133–44 (2010).

18. P. R. Evans, An introduction to data reduction: space-group determination, scaling and intensity statistics, *Acta Crystallographica. Section D, Biological crystallography* **67**, 282–292 (2011).

19. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallographica. Section D, Biological crystallography* **66**, 213–221 (2010).

20. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot., *Acta Crystallographica. Section D, Biological crystallography* **66**, 486–501 (2010).

21. F. Sobott, H. Hernández, M. McCammon, M. Tito, C. Robinson, A Tandem Mass Spectrometer for Improved Transmission and Analysis of Large Macromolecular Assemblies, *Anal. Chem.* **74**, 1402–1407 (2002).

22. H. Hernandez, C. Robinson, Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry, *Nat. Protoc.* **2**, 715–726 (2007).

23. D. I. Svergun, Determination of the regularization parameter in indirect-transform methods using perceptual criteria, *J. Appl. Crystallogr.* **25**, 495–503 (1992).

24. A. Sali, T. L. Blundell, Comparative protein modelling by satisfaction of spatial restraints., *J. Mol. Biol.* **234**, 779–815 (1993).

25. D. Svergun, C. Barberato, M. H. J. Koch, *CRYSOL* – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates, *J. Appl. Crystallogr.* **28**, 768–773 (1995).

26. D. I. Svergun, Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing, *Biophys. J.* **76**, 2879–2886 (1999).

27. M. B. Kozin, D. I. Svergun, Automated matching of high- and low-resolution structural models, *J. Appl. Crystallogr.* **34**, 33–41 (2001).

28. V. V. Volkov, D. I. Svergun, Uniqueness of *ab initio* shape determination in small-angle scattering}, *J. Appl. Crystallogr.* **36**, 860–864 (2003).

29. E. F. Pettersen *et al.*, UCSF Chimera–a visualization system for exploratory research and analysis., *J. Comput. Chem.* **25**, 1605–1612 (2004).

30. S. Britton, J. Coates, S. Jackson, A new method for high-resolution imaging of Ku foci to decipher mechanisms of DNA double-strand break repair, *J. Cell Biol.* **202**, 579–595 (2013).

31. M. Stucki *et al.*, MDC1 Directly Binds Phosphorylated Histone H2AX to Regulate Cellular Responses to DNA Double-Strand Breaks, *Cell* **123**, 1213–1226 (2005).

32. P. Hsu *et al.*, DNA targeting specificity of RNA-guided Cas9 nucleases, *Nat. Biotech.* **31**, 827–832 (2013).

33. P. Mali *et al.*, RNA-Guided Human Genome Engineering via Cas9, *Science* **339**, 823–826 (2013).

34. S. Polo *et al.*, Regulation of DNA-End Resection by hnRNPU-like Proteins Promotes DNA Double-Strand Break Signaling and Repair, *Mol. Cell* **45**, 505–516 (2012).

35. K. Mizuguchi, C. M. Deane, T. L. Blundell, M. S. Johnson, J. P. Overington, JOY: protein sequence-structure representation and analysis., *Bioinformatics* **14**, 617–623 (1998).

36. A. Innis, J. Shi, T. Blundell, Evolutionary trace analysis of TGF-β and related growth factors: implications for site-directed mutagenesis, *Protein Eng.* **13**, 839–847 (2000).

37. S. Altschul *et al.*, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* **25**, 3389–3402 (1997).

38. J. Shi, T. Blundell, K. Mizuguchi, FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties., *J. Mol. Biol.* **310**, 243–257 (2001).
39. L. A. Kelley, M. J. Sternberg, Protein structure prediction on the Web: a case study using the Phyre server., *Nat. Protoc.* **4**, 363–371 (2009).